Unfolding spatiotemporal representations of 3D visual perception in the human brain

Zitong Lu (lu.2637@osu.edu) and Julie D. Golomb (golomb.9@osu.edu)

Department of Psychology, The Ohio State University

Abstract

Although visual input is initially recorded in two dimensions on our retinas, we perceive and interact with the world in three dimensions. Achieving 3D perception requires the brain to integrate 2D spatial representations with multiple depth cues, such as binocular disparity. However, most studies typically examine 2D and depth information in isolation, leaving the integrated nature of 3D spatial encoding largely underexplored. In this study, we collected a large-scale, multimodal neuroimaging dataset consisting of multiple EEG and fMRI sessions while participants viewed stereoscopic 3D stimuli through red-green anaglyph glasses. Participants first completed a behavioral session including depth judgement tasks and a novel cube adjustment task to quantify and calibrate individual depth perception in units of binocular disparity. Then during two EEG and two fMRI sessions, participants passively viewed stimuli presented at 64 systematically sampled 3D locations, yielding over 66,000 trials in total across ten participants. Combining this large-scale EEG-fMRI dataset with computational methods via representational similarity analysis, we not only systematically characterized the spatiotemporal representation of multiple spatial features in 3D perception but also explored how different coordinate systems (e.g., Cartesian or Polar) might be employed across brain regions and time. Our results reveal that human brains employ multiple types of spatial feature representations and coordinate systems to encode spatial locations at different temporal stages and across distinct cortical regions. In addition to strong representations of 2D space throughout visual cortex, we find unique representations for depth and 3D features in later timepoints and visual areas, including some evidence for 3D processing in parahippocampus. These findings contribute to a more comprehensive understanding of the spatiotemporal organization of neural representations that support 3D perception. Additionally, our novel large dataset will be made openly available to support future research on 3D perception and spatial cognition.

Key words: 3D perception, EEG, fMRI, coordinate system, spatial location

Introduction

Despite the inherently two-dimensional nature of retinal input, humans experience the world as a richly structured three-dimensional environment. The human visual system has the ability to integrate 2D spatial representations with a range of depth cues– including binocular disparity, perspective, shading, relatize size, occlusion, and motion – to construct rich 3D spatial percepts that support perception, navigation, and interaction with the environment (Howard, 2012; Welchman, 2016). Yet despite its centrality, the neural mechanisms that enable this transformation from 2D input to 3D visual perception remain poorly understood. A fundamental theoretical question in vision science is how the brain builds internal representations of 3D space: What spatial features and coordinate systems are used to encode location? How are depth and 2D features integrated into unified percepts? And how do these representations evolve across cortical regions and time? Addressing these questions is critical for understanding not only the computations that underlie visual perception, but also broader principles of neural coding, spatial cognition, and multisensory integration.

Decades of research have investigated how the brain encodes 2D spatial dimensions of space (T. Carlson et al., 2011; Engel et al., 1994; Fischer et al., 2011; Golomb & Kanwisher, 2012; Grill-Spector & Malach, 2004; Kravitz et al., 2010; Maunsell & Newsome, 1987; Schwarzlose et al., 2008; Sereno et al., 1995; Silver & Kastner, 2009; Tootell, Hadjikhani, Mendola, et al., 1998; Wandell et al., 2007), as well as how visual areas respond to depth information conveyed by binocular disparity and other cues (Backus et al., 2001; Bridge et al., 2023; Chen et al., 2020; Ip et al., 2014; Neri et al., 2004; Uka & DeAngelis, 2006). Recent findings further suggest a gradual transition from 2D to depth representations along the visual hierarchy (Finlayson et al., 2017; Henderson et al., 2019). However, most prior work has investigated or analyzed 2D and depth in isolation, leaving the integrated nature of 3D visual perception – how individual spatial features are jointly represented and how these features converge into coherent 3D representations – largely unexplored.

An additional layer of complexity stems from the fact that spatial locations can be encoded in multiple coordinate systems. In 3D space, Cartesian, Polar, Cylindrical, and Spherical coordinate systems are all geometrically valid and may serve distinct computational functions across brain regions. Previous studies using fMRI retinotopic mapping, population receptive field (pRF) modeling, or multivariate pattern analysis have demonstrated cortical encoding of polar-based features such as eccentricity and polar angle (Carvalho et al., 2020; Conner et al., 2004; Silver & Kastner, 2009; Tootell et al., 1997; Tootell, Hadjikhani, Vanduffel, et al., 1998; Tu et al., 2022; Wandell et al., 2007; Warnking et al., 2002; Wilkinson et al., 2000), cartesian-based features such as horizontal and vertical position (J. M. Carlson et al., 2011; Fischer et al., 2011; Golomb & Kanwisher, 2012; Graumann et al., 2022; Kravitz et al., 2010; Lu et al., 2022; Schwarzlose et al., 2008), and position-in-depth features (Alvarez et al., 2021; Bridge & Parker, 2007; Finlayson et al., 2017; Golomb, 2018; Henderson et al., 2019; Ip et al., 2014). However, no prior work has comphrehensively examined the brain's representation of higher-dimensional spatial features

such as 3D radius or 3D polar angle, nor systematically compared the contributions of different coordinate systems within the same 3D stimulus space. It remains unknown whether the brain uses multiple coordinate systems in parallel, whether specific systems dominate in particular brain regions or time windows, and how such representations emerge and evolve across space and time.

Three major methodological challenges have limited progress on these questions. First, perceived position-in-depth can vary across individuals and depth cues, and it is not straightforward to assume that a fixed distance in binocular disparity corresponds to the same perceived distance in depth across individuals. Most depth-based neuroimaging studies use fixed disparity levels across participants, precluding accurate quantitative comparisons between depth and horizontal/vertical dimensions (Alvarez et al., 2021; Bridge & Parker, 2007; Finlayson et al., 2017). Second, high-resolution characterization of 3D encoding demands large-scale neural datasets to precisely measure feature-level and coordinate-level brain representations. Although recent advances in large-scale fMRI and EEG datasets have opened new possibilities for investigating the neural basis of visual cognition (Allen et al., 2022; Chang et al., 2019; Gifford et al., 2022; Grootswagers et al., 2022; Hebart et al., 2023), no such dataset currently exists for 3D spatial perception. Third, most neuroimaging studies in humans have focused solely on depth encoding using fMRI, without characterizing temporal dynamics, leaving the spatiotemporal structure of 3D visual perception largely uncharted.

To overcome these challenges and deeply probe the neural underpinnings of 3D perception, we developed a multimodal framework for tracking 3D spatial representations across time and brain space. First, to precisely quantify perceived depth from binocular disparity, all participants performed an initial behavioral “cube adjustment” task in which participants dynamically adjusted the distances between 3D stimuli to match perceived horizontal, vertical, and depth distances; i.e., until the stimuli formed the corners of a perfect cube (Figure 1A). This individualized calibration allowed us to derive a parametric mapping between physical disparity and perceived 3D depth for each participant, such that we could select an appropriately calibrated stimulus space for each participant ensuring accurate reconstruction of their subjective 3D spatial experience. Then, we collected an extensive dataset of 3D visual perception comprising both EEG and fMRI from multiple sessions per participant (Figure 1A). Stimuli for these neuroimaging sessions were viewed stereoscopically through red-green anaglyph glasses and sampled a structured 3D space with 64 distinct spatial locations (4×4×4 grid centered on fixation). For each individual, we calculated the neural representational similarity between each pair of stimuli at each point in time (every 20 ms) from the EEG data, and each point in brain space (voxel-by-voxel searchlights or functional regions of interest) from the fMRI data. Because each stimulus location could be defined in multiple potential coordinate systems and in terms of multiple spatial features, we could compare the neural patterns to different hypothetical representational patterns (Figure 1B-E; also see more details in Methods) to characterize how 3D spatial location is represented across multiple coordinate systems in both time and brain space.

By integrating individualized depth calibration, large-scale EEG and fMRI recordings, representational similarity analysis (RSA), and an expanded set of 3D spatial features encompassing both Cartesian and Polar coordinate systems, our unique, publicly available dataset enables a fine-grained analysis of how the human brain encodes 2D, depth, and integrated 3D information. This allows us to uncover the spatiotemporal dynamics of multiple spatial features and coordinate systems, highlighting how the brain flexibly transforms spatial representations in complex 3D environments.

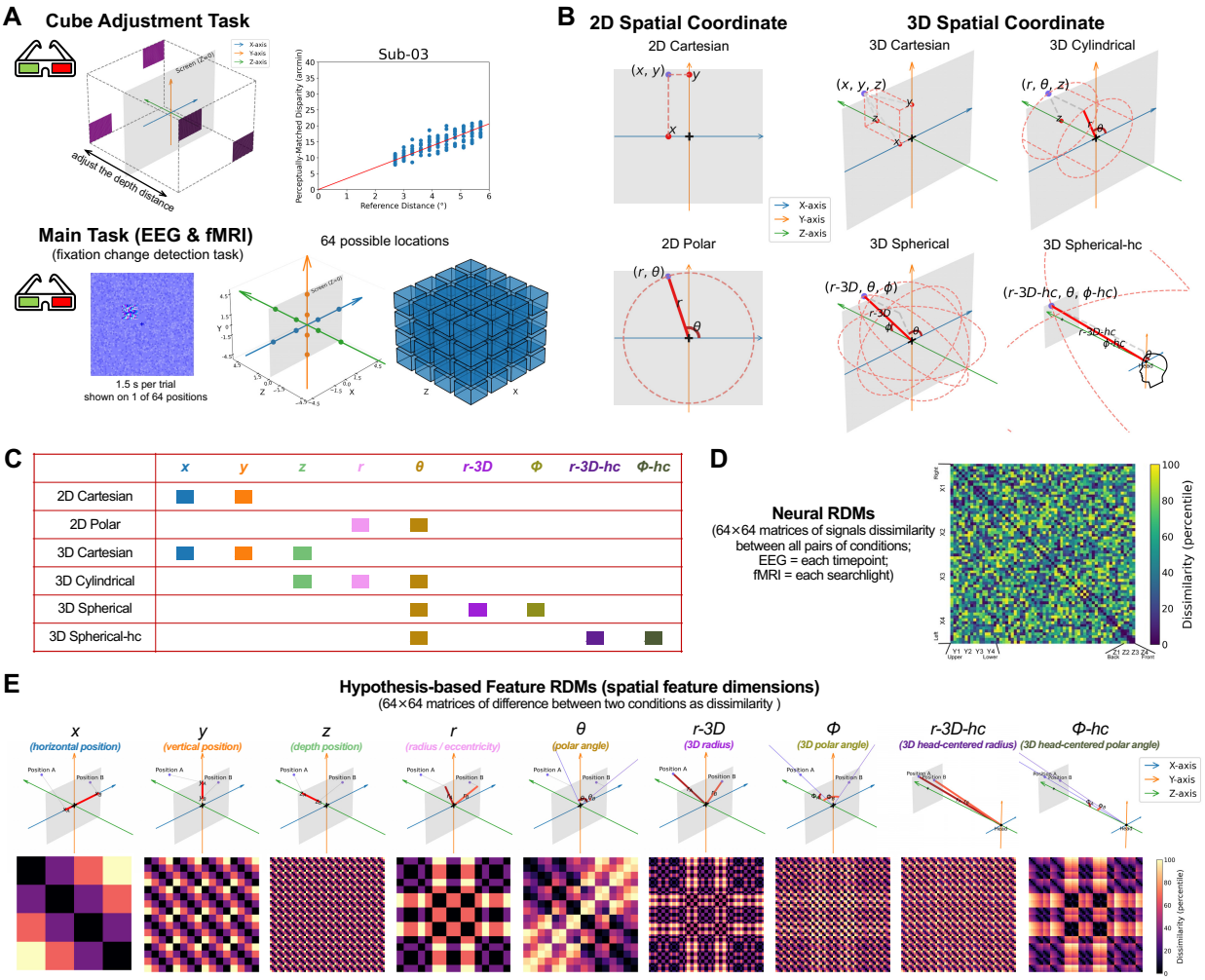

Figure 1 Experimental design, coordinate systems, and representational patterns.
(A) Tasks in our study. In Cube Adjustment Task for individualized depth calibration, participants adjusted the depth distance between the front and back panels using the vertices of the stimuli to generate a virtual cube, allowing us to derive a personalized mapping between binocular disparity and perceived depth distance. Upper-left: Schematic of the cube adjustment task. Upper-right: Example calibration function fitted based on one subject's (Sub-03) real behavioral data. In Main EEG and fMRI task, participants performed a fixation-change detection task while passively viewing stimuli presented at one of 64 3D spatial locations arranged in a 4×4×4 grid. Stimuli were small cubes of high-contrast dynamic random dot stimuli presented on a lower-contrast static background viewed through red-green anaglyph

glasses. (B) Schematics illustrating the possible 2D and 3D spatial coordinate systems used to define spatial features. For each schematic, the black fixation cross indicates the fixation location and the gray dot the example stimulus location, with the light gray rectangle depicting the computer monitor. (C) Table showing the spatial coordinate systems (rows) and which spatial feature dimensions (columns) they contain. (D) Neural RDMs (64 x 64 matrices) were computed separately for each EEG time point or each fMRI searchlight, capturing pairwise signal dissimilarity between the 64 spatial conditions. (E) Hypothesis-based feature RDMs were constructed for each of the nine different spatial feature dimensions by calculating pairwise differences in feature values between conditions. Top row shows geometric representations of an example pair of stimuli (Position A and Position B) and how dissimilarity is calculated for each spatial feature dimension (red lines; e.g. comparison of x_A and x_B). Bottom row shows the 64x64 RDM for each spatial feature dimension. Neural RDMs were compared with Hypothesis-based RDMs via representational similarity analysis.

Results

In our study, we aimed to uncover spatiotemporal neural representations of different spatial dimensions and coordinates comprising human 3D perception. We collected a large-scale EEG-fMRI dataset of healthy young adults viewing 64 locations sampled across individually perceptually-calibrated 3D space (Figure 1A).

To characterize how different 3D spatial features and coordinates (Figure 1B-C) are represented in the human brain over time and across brain space, we applied representational similarity analysis (RSA). Neural representational dissimilarity matrices (neural RDMs; Figure 1D) were computed for each time point using EEG data and for each searchlight location or functional region of interest (ROI) using fMRI data. In parallel, we constructed nine hypothesis-based feature RDMs (Figure 1E) corresponding to nine possible spatial features – horizontal position (x), vertical position (y), depth position (z), radius / eccentricity (r), polar angle (θ), 3D radius (*r-3D*), 3D polar angle (Φ), 3D head-centered radius (*r-3D-hc*), and 3D head-centered polar angle (Φ*-hc*) from six possible spatial coordinate systems (2D Cartesian, 2D Polar; 3D Cartesian, 3D Cylindrical, 3D Spherical, and 3D Spherical-hc). To isolate the unique contribution of each feature, we computed partial correlations between neural RDMs and individual hypothesis-based feature RDMs, controlling for all other hypothesis-based RDMs. This approach allowed us to identify the distinct spatiotemporal representations of each spatial feature, controlling for variance explained by other features. Moreover, because each participant viewed an identical set of stimuli twice under different binocular disparity settings (flipping the direction of the red-green anaglyph glasses between sessions), our analyses reflect perceived depth from disparity, controlling for any monocular-based cues or potential low-level color or contrast differences across eyes.

Representations of 2D and 3D spatial features

First, to establish the overall spatiotemporal organization of spatial encoding of 2D, depth, and 3D spatial features in the human brain, we grouped the spatial features into three categories: 2D

spatial features (x, y, r, θ), position-in-depth feature (z), and 3D spatial features (*r-3D*, ϕ, *r-3D-hc*, ϕ*-hc*). Here, representational similarity was averaged across all features within each category at each EEG time-point and each fMRI searchlight unit, yielding temporal and spatial profiles for 2D, depth, and 3D encoding. As shown in Figure 2A, EEG results revealed a clear temporal hierarchy from 2D to depth: 2D features were represented earliest and strongest (40-1800 ms), followed by depth features, which showed relatively weaker but temporally extended encoding – first emerging transiently in an early stage (120-280 ms) and reappearing more robustly in a late stage (620-1800 ms). However, we did not observe a significant time-window of 3D feature representation when the four 3D spatial features were averaged together. fMRI searchlight results (Figure 2B; see also Figure S1) demonstrated a complementary spatial hierarchy. 2D representations (red) were broadly distributed across early visual cortex, posterior parietal cortex, and ventral temporal cortex. Depth encoding (green) appeared in spatially scattered patches, with some partial overlap observed in both temporal and parietal regions. 3D feature representations (blue) yielded only two significant clusters located in the left middle occipital and temporal regions. This weak representation of 3D feature in both EEG and fMRI is likely due to their temporal and spatial heterogeneity – different 3D spatial features might be encoded at distinct times and in different brain regions. Supporting this, Figure S2 shows the number of significant 2D and 3D features over time and brain space, indicating that 3D features tend to emerge later than 2D and depth features and are distributed across a broader range of cortical areas, from posterior to anterior, albeit with lower consistency. Together, these results highlight a spatiotemporal gradient in the encoding of spatial features – across time, progressing from early 2D to later depth and 3D representations; and across space, shifting from occipital to frontal regions, with 2D features dominant posteriorly and depth and 3D features engaging more anterior areas.

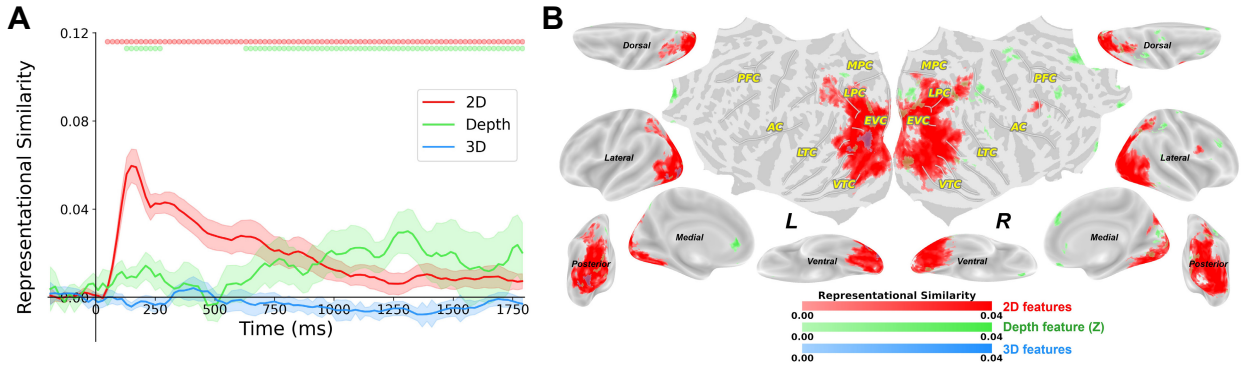

Figure 2 Grouped spatiotemporal representation of 2D, depth, and 3D spatial features.
(A) EEG time courses of representational similarity, averaged across 2D (x, y, r, θ), depth (z), and 3D (*r-3D*, ϕ, *r-3D-hc*, ϕ*-hc*) features. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$). (B) fMRI searchlight results highlighting the cortical distribution of 2D, depth, and 3D representations. Colored voxels indicate the similarity averaged among grouped features for the searchlight centered on that voxel was significant (permutation test, cluster-corrected, $p < .01$).

Spatiotemporal representations of individual spatial features

Then we further visualized the spatiotemporal representations of individual spatial features in time and brain space. To dissect the temporal dynamics of the 9 individual spatial features, Figure 3A shows the EEG RSA time courses of representation strength in detail, grouping the features according to their coordinate system(s). To quantify the timing of representational emergence and peak, we calculated the significant time windows for each feature dimension (Figure 3A), and extracted the onset and peak latencies for each feature at the individual-subject level (Figure 3B). Onset latency was defined as the first timepoint at which representational similarity was significant, and peak latency was defined as the timepoint at which the maximal partial correlation occurred. Additionally, we averaged representational similarity values within three broad time windows: 0-240 ms, 240-740 ms, and 740-1800 ms (Figure 3C). During the early window (0-240 ms), 2D features such as x, y, and θ exhibited significant representational similarity. In the middle window (240-740 ms), x and θ remained significant, while r and 3D Φ features also exhibited significant representational similarity. In the late window (740-1800 ms), the position-in-depth z feature became significant, along with persisting r and θ, but no longer x and y features.

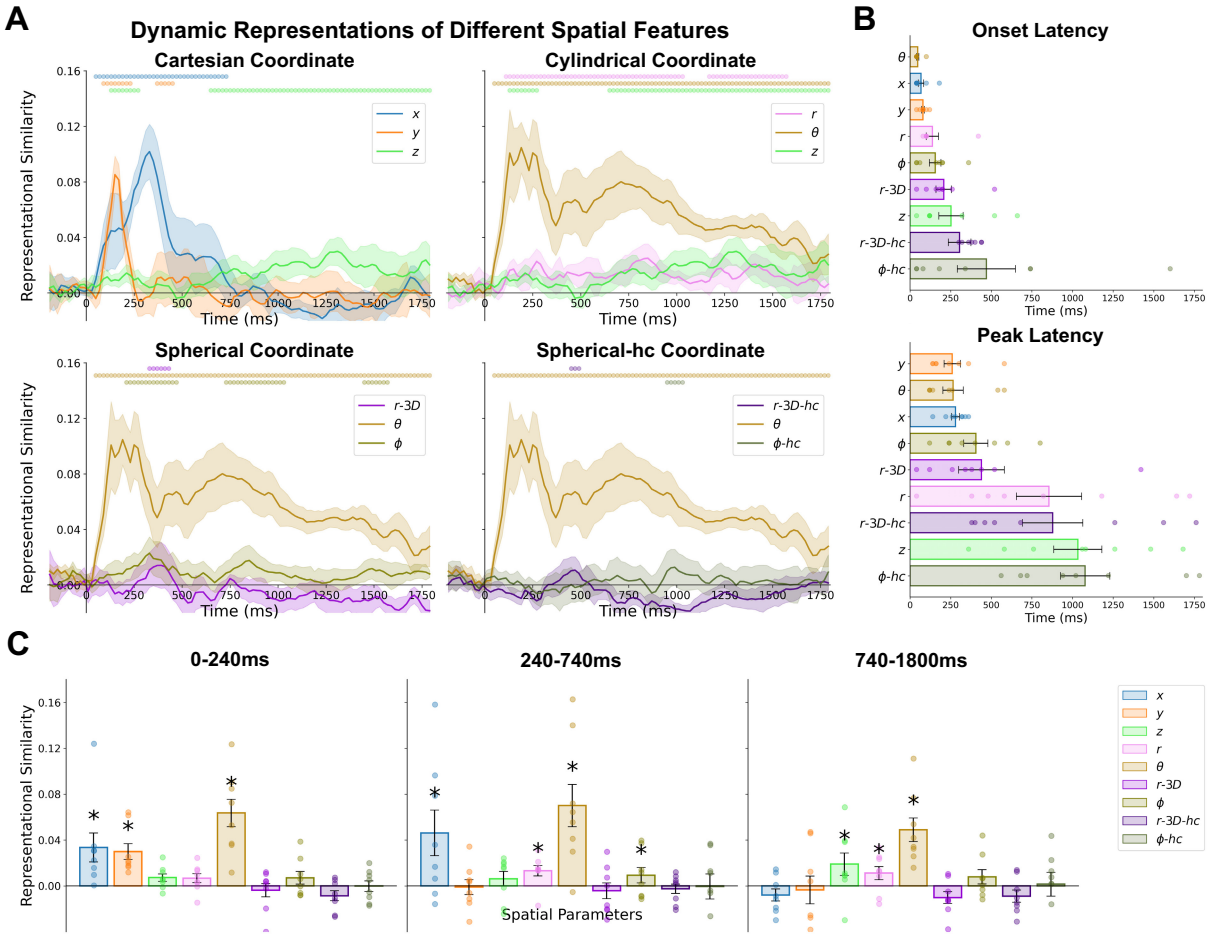

Figure 3 Dynamic encoding of spatial features in EEG.

(A) Time-resolved representational similarity (partial Spearman correlation) between EEG temporal RDMs and the nine hypothesis-based spatial feature RDMs, grouped by spatial coordinate systems. Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$). (B) Onset and peak latencies of each spatial feature's EEG representational similarity result. Dots denote individual participants; error bars indicate SEM. (C) Averaged representational similarity in three time windows (0–240 ms, 240–740 ms, 740–1800 ms). Dots denote individual participants; error bars indicate SEM; asterisks denote features with significant group-level similarity (permutation test, $p < .01$).

Figure 4A shows the parallel fMRI RSA analysis, revealing partially overlapping but dissociable cortical topographies of individual spatial feature representations. The 2D features – *x*, *y*, r, and θ – were most strongly represented in traditional visual cortices, including early visual occipital cortex, ventral temporal cortex, and posterior parietal cortex, though *x* and *y* representations in particular seemed to extend further along both the ventral and dorsal visual pathways. In contrast, the position-in-depth *z* feature showed more restricted activation patterns. Most of the 3D features, such as *r-3D*, *r-3D-hc*, and ϕ*-hc,* exhibited distributed representations more heavily focused in frontal and parietal regions, though ϕ representations were more focused in occipital visual cortex.

We also quantified these representational similarity patterns within predefined functional and anatomical ROIs (Figure 4B). Early visual areas (V1–V4) showed robust encoding of 2D *x*, *y*, *r*, and θ features ($p < 0.01$, permutation test, cluster-based corrected), consistent with their known 2D retinotopic organization. Depth feature, *z*, was significantly represented in V3b and VO1, areas previously associated with perception of depth information. 3D features such as *r-3D* were significantly represented in dorsal areas including V3v, V3a, and IPS1-5. *r-3D-hc* were significantly represented in V3v and IPS0. Φ was significantly represented in V2d, V3d, V3a, V4, LO1, LO2, and MT. And Φ*-hc* was significantly represented in V3a, V3b, VO1, VO2, and FEF. Notably, medial temporal lobe structures such as the hippocampus, parahippocampal gyrus, and entorhinal cortex showed no significant representation of any individual spatial features.

For additional exploration and hypothesis-testing beyond these pre-defined ROIs, our full dataset and code are openly available and easily integrable with various atlases and parcellations. Combined visualization of EEG and fMRI results are also illustrated in Supplementary Video 1. For the remainder of this paper, we now turn to additional theoretically-driven questions about 3D spatial representation and integration enabled by this approach.

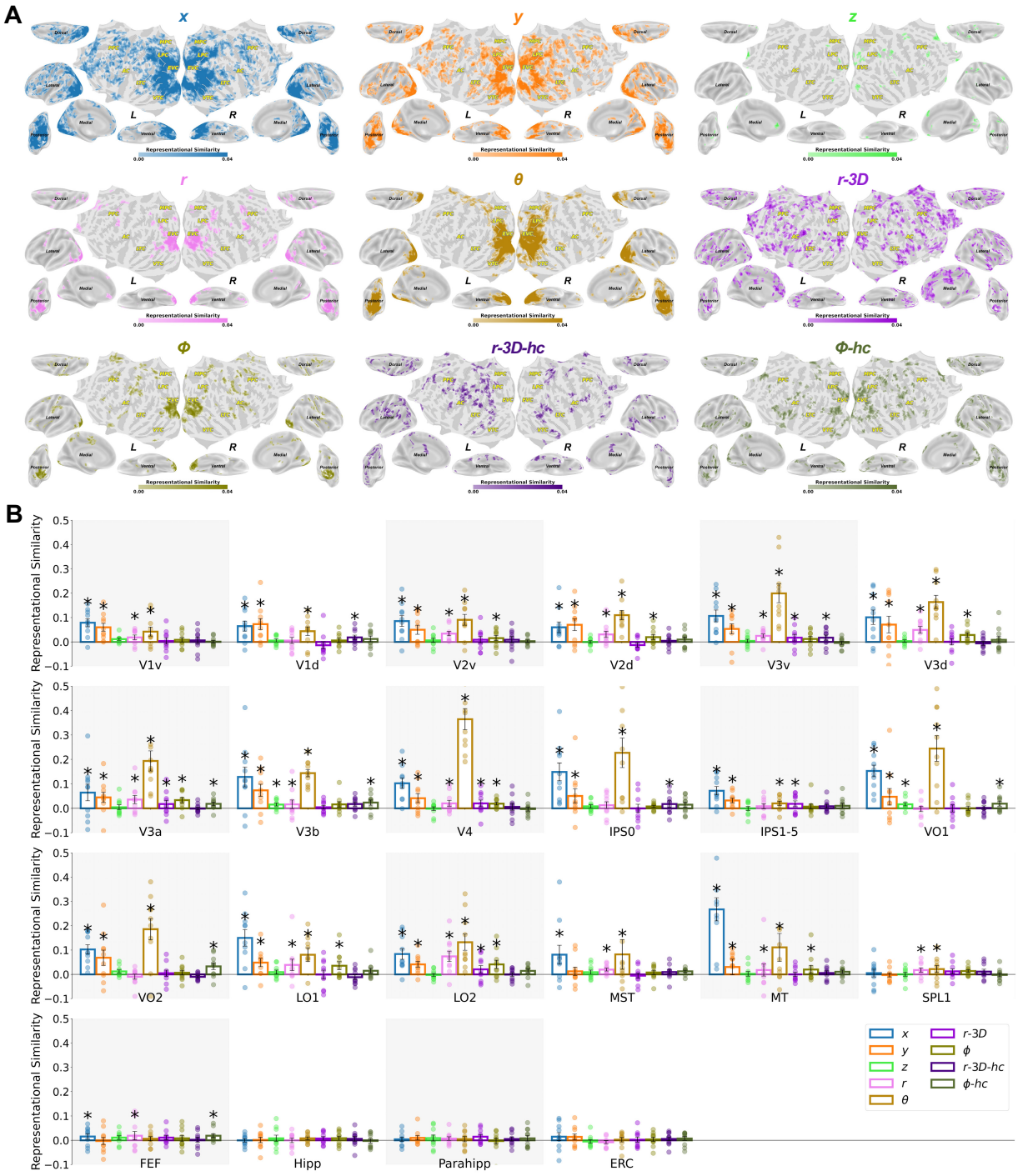

Figure 4 Spatial encoding of spatial features in fMRI.
(A) fMRI searchlight representational similarity (partial Spearman correlation) results for each spatial feature, displayed on inflated cortical surfaces. Colored voxels indicate the searchlight centered on that voxel was significant (permutation test, cluster-based corrected, $p < .01$). (B) ROI-based RSA results showing group-level representational similarity across 25 predefined ROIs. Asterisks indicate significant feature representations (permutation test, $p < 0.01$). Dots denote individual participants. Error bars indicate SEM.

Are different spatial coordinate systems preferentially represented across time and brain space?

Building on the feature-level analysis, we next examined if the human brain is representing visual space in a particular spatial coordinate system(s) at different points in time or brain space. We defined a coordinate system as completely represented if all individual constituent features within that coordinate system were significantly represented.

For the 2D coordinate systems, EEG-based results revealed early complete representations of 2D Cartesian coordinate system and more sustained complete representations of 2D Polar coordinate system (Figure 5A), with the Polar representation significantly stronger than Cartesian in the later time window (Figure 5B). The fMRI-based results (Figure 5C) revealed significant evidence of both 2D Cartesian and Polar representations in classic visual areas, but strikingly, the cortical topographies were largely dissociated. Polar representations were more spatially confined in early visual cortex and more foveal areas, with the Cartesian representations showing more widespread activation into peripheral visual areas and across higher-level dorsal and ventral visual pathway regions, extending into later temporal, parietal, and some prefrontal areas.

In terms of the 3D coordinate systems, these representations emerge later than the 2D representations in time, and with much sparser cortical coverage. With EEG we found brief complete representations of 3D Cartesian and 3D Spherical coordinate systems, and more sustained complete representations of 3D Cylindrical (Figure 5A-B). With fMRI, complete 3D coordinate representations were far more restricted. While the individual spatial features were broadly distributed, only small patches of cortex showed significant encoding of all three features in 3D Cartesian (x, y, and z) and 3D Spherical (r-$3D$, θ, Φ) coordinate systems, and no voxels were significant for the other two 3D complete coordinate systems (Figure 5C). These findings suggest that although multiple 3D coordinate systems may be partially engaged in a given region, only a subset exhibit spatially overlapping or temporally stable complete encoding of all constituent spatial features during perception. As a supplementary analysis for these four different 3D coordinate systems, we relaxed the criterion for complete representations to examine emergent representations – defined as the presence of at least one significantly encoded coordinate-specific feature (i.e., a feature uniquely associated with that coordinate system – e.g., r is specific to the 3D Cylindrical coordinate, while x and y are unique to the 3D Cartesian system) from each 3D coordinate system, and found more widespread representations with consistent regional preferences (Figure S3).

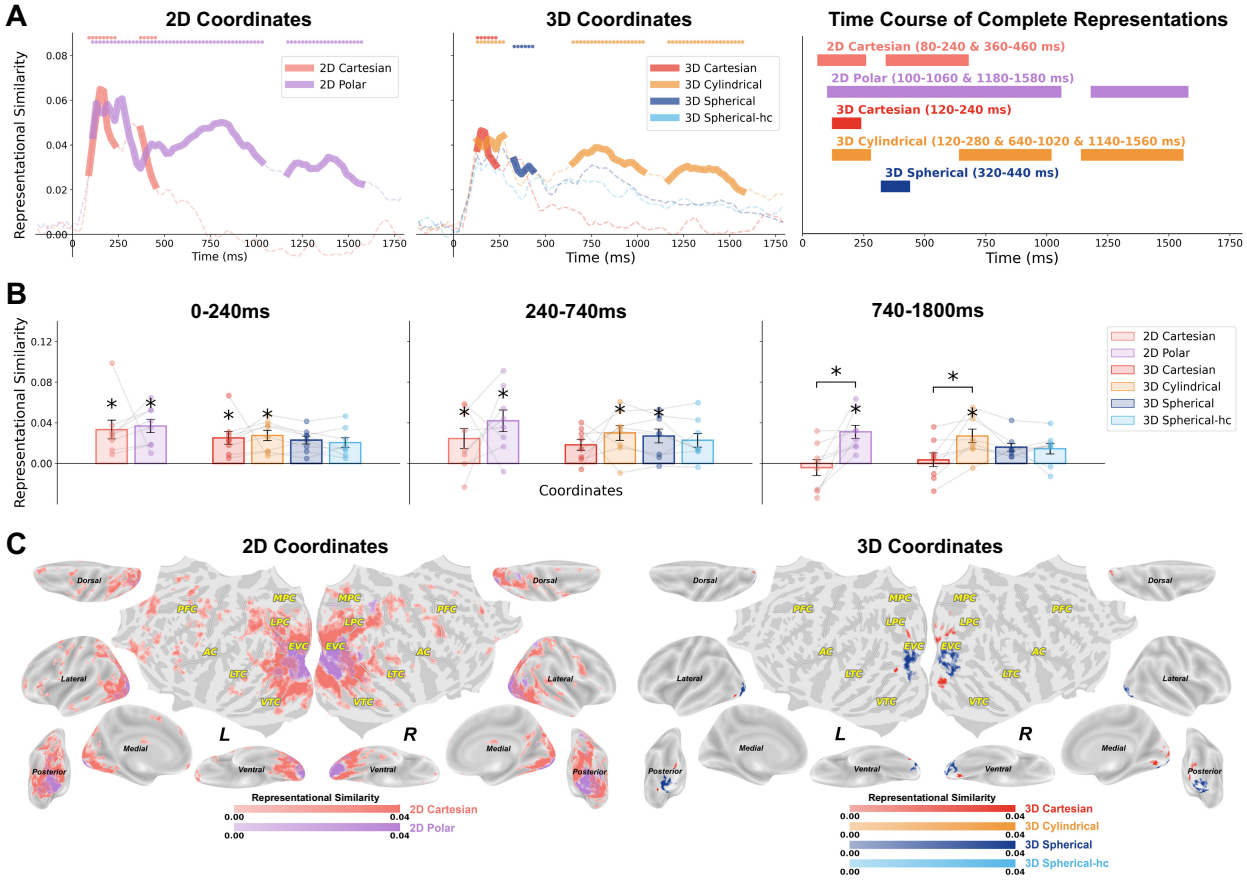

Figure 5 Spatiotemporal representations of complete coordinate systems (all feature components within a coordinate system significant). (A-B) EEG Representational similarity results. (A) Time courses of significant complete representations of 2D and 3D coordinates. Thinkened lines and Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$). (B) EEG representational similarity for each complete coordinate system averaged over three time windows (0–240 ms, 240–740 ms, 740–1800 ms). Dots denote individual participants. Asterisks indicate significance (permutation test, $p < 0.01$); error bars indicate SEM. (C) fMRI searchlight maps of complete coordinate representations. Voxels were labeled as significant if they showed representational similarity for all features in a certain coordinate system (permutation test, cluster-based corrected, $p < .01$).

Does the human brain also represent spatial locations through integrated 2D and 3D geometric distances?

While the preceding analyses focused on the encoding of individual spatial features and their combinations, it remains unclear whether the brain also processes spatial location in an integrated, geometric manner – that is, whether it could form a unified 2D or 3D positional coding pattern which reflects not the individual feature dimensions but their combined geometric relationships in space. For example, two objects that differ by one unit horizontally and one unit

vertically would be separated by $\sqrt{2}$ units in an integrated 2D Euclidean space. A feature-based encoding scheme might treat these two objects as distinct along independent horizontal and vertical maps, while a geometric distance representation would capture their spatial dissimilarity as a function of their combined distance in this space. Critically, such geometric distance representations highlight a fundamentally different kind of spatial information. While feature-based maps may support tasks like localizing and reaching toward a single object, geometric-distance-based codes may underlie the perception of global spatial structure and inter-object relationships.

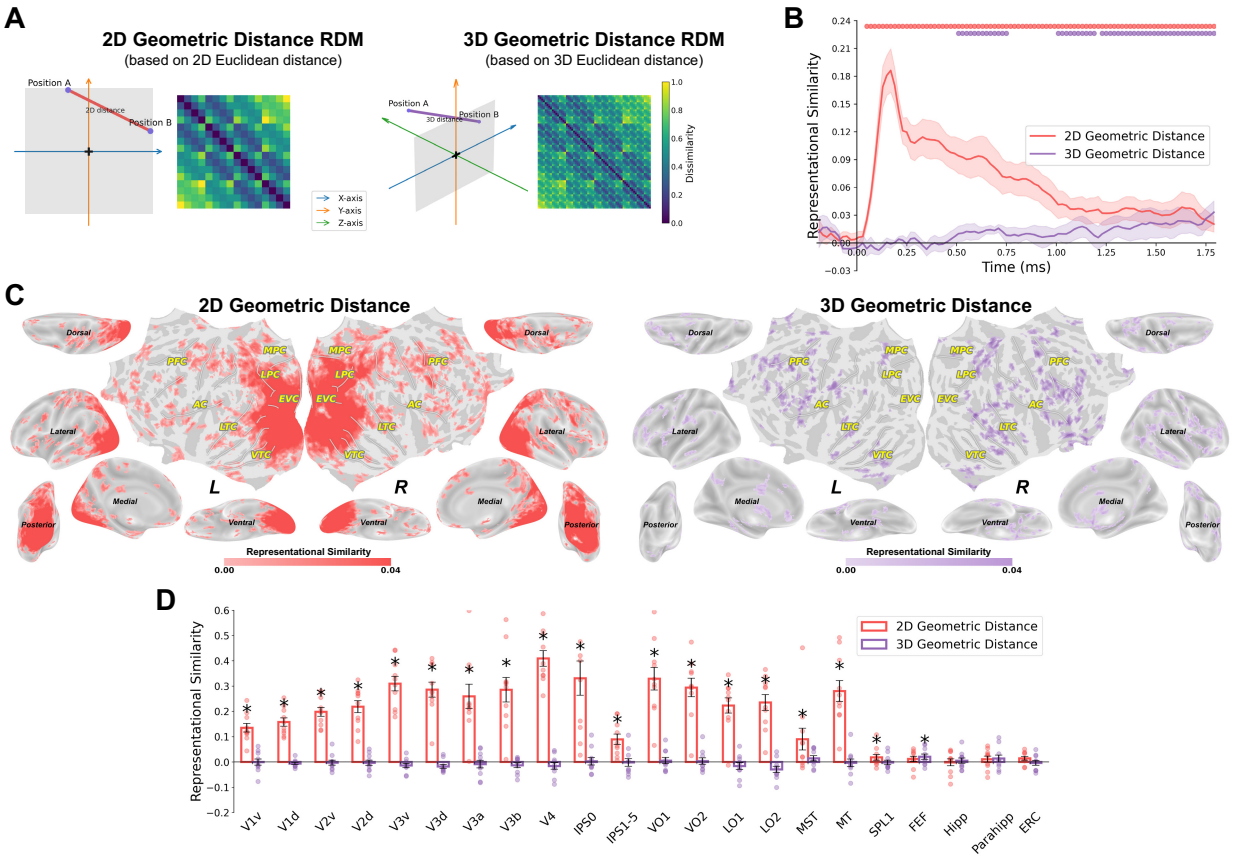

Figure 6 Spatiotemporal representations of 2D and 3D geometric distance.
(A) Two geometric distance RDMs capturing holistic spatial location similarity based on 2D and 3D Euclidean distances across conditions. Top row shows geometric representations of an example pair of stimuli (Position A and Position B) and how dissimilarity is calculated based on the Euclidean distance between two positions in 2D or 3D space. Bottom row shows the 64x64 RDM for each geometric distance. (B) Time-resolved representational similarity (partial Spearman correlation) between EEG temporal RDMs and the two geometric distance RDMs. Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$). (C) fMRI searchlight representational similarity (partial Spearman correlation) results for 2D and 3D geometric distances, displayed on inflated cortical surfaces. Colored voxels indicate the searchlight

centered on that voxel was significant (permutation test, cluster-based corrected, $p < .01$). (D) ROI-based RSA results showing group-level representational similarity of 2D and 3D geometric distances across 25 predefined ROIs. Asterisks indicate significant feature representations (permutation test, $p < 0.01$). Dots denote individual participants. Error bars indicate SEM.

To test this, we constructed two geometric distance RDMs (Figure 6A): a 2D geometric distance RDM based on pairwise Euclidean distances in the 2D plane, and a 3D geometric distance RDM based on Euclidean distanced in full 3D space. These geometric distance RDMs capture the similarity structure that would be expected if the human brain encoded locations holistically rather than feature-by-feature. We first examined the extent to which neural patterns tracked these 2D and 3D geometric distance models, computing partial correlations between the neural RDMs and each hypothesis-based geometric distance RDM while controlling for the other (2D or 3D). As shown in Figure 6B, EEG analyses revealed robust encoding of both 2D and 3D geometric distances. Notably, 2D geometric distance representations emerged rapidly and remained significant throughout the epoch (40-1800 ms), and 3D geometric distance representations emerged later (480-740, 980-1180, and 1200-1800 ms) and were weaker than 2D geometric distance representations. fMRI analyses (Figure 6C-D) showed broad cortical representation for 2D geometric distance across the visual cortex. In contrast, 3D geometric distance representations were more spatially restricted and sparser, localized to patches in parietal, lateral occipital, and prefrontal regions, such as FEF. This spatiotemporal dissociation suggests a possible progression from planar to volumetric spatial integration in visual processing, analogous to the progression from 2D to 3D individual spatial feature dimensions.

While these initial geometric distance analyses revealed robust 2D to 3D representations, they may still partially reflect variance attributable to individual spatial features, given the possible collinearity between geometric distances and feature dimensions. To further isolate the unique contributions of these geometric distance representations beyond feature-level coding, we repeated the analysis above while additionally controlling for all nine individual spatial feature RDMs. This approach removes variance shared with feature-based coordinate representations, isolating the residual geometric integration component that may be otherwise masked by spatial features. EEG analyses (Figure 7A) revealed that unique 2D geometric distance representations persisted from early to late stages (100-240, 320-960, and 1320-1680 ms), whereas uniqe 3D geometric distance representations emerged only later (620-900 and 940-1160 ms). fMRI analyses (Figure 7B and 7C) revealed widespread unique 2D geometric distance representations across the visual cortex. However, 3D integration was highly sparse, localized to small significant clusters in precentral gyrus, superior temporal gyrus, caudate, and the parahippocampal cortex – a region known to be involved in spatial navigation and scene representation (Aminoff et al., 2013; R. Epstein et al., 1999, 2003; R. A. Epstein, 2008).

Together, these findings demonstrate that the human brain encodes spatial locations not only via independent features or coordinate systems but also through holistic geometric distance codes.

Moreover, 2D geometric integration emerges earlier, stronger, and more widely distributed, whereas 3D geometric integration emerges later, weaker, and in more sparsely localized higher-order regions.

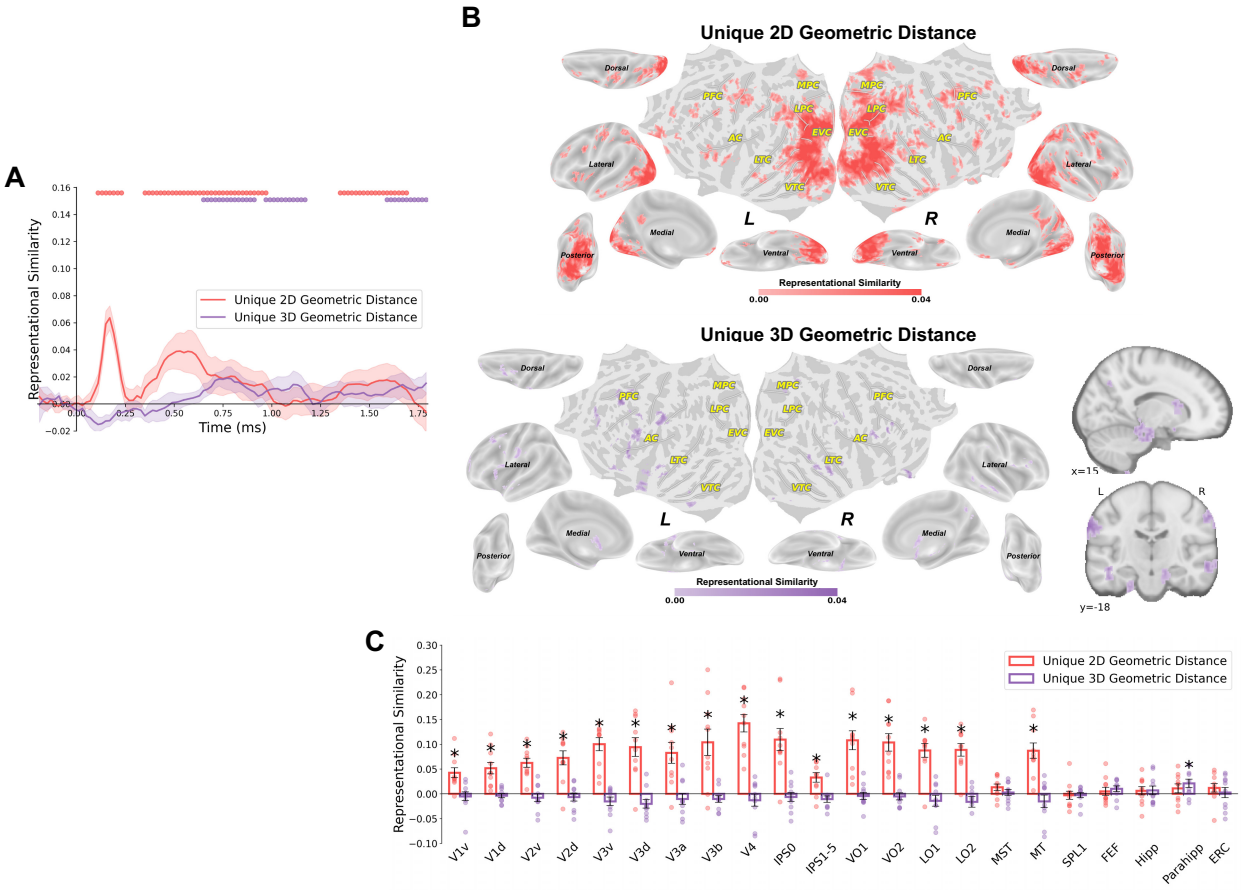

Figure 7 Unique spatiotemporal representations of 2D and 3D geometric distance beyond spatial feature coding.

(A) Time-resolved representational similarity (partial Spearman correlation, also controlling nine feature RDMs) between EEG temporal RDMs and the two geometric distance RDMs. Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$). (C) fMRI searchlight representational similarity (partial Spearman correlation, also controlling nine feature RDMs) results for unique 2D and 3D geometric distances, displayed on inflated cortical surfaces. Colored voxels indicate the searchlight centered on that voxel was significant (permutation test, cluster-based corrected, $p < .01$). (D) ROI-based RSA results showing group-level representational similarity of unique 2D and 3D geometric distances across 25 predefined ROIs. Asterisks indicate significant feature representations (permutation test, $p < .01$). Dots denote individual participants. Error bars indicate SEM.

Are individual differences in depth perception reflected in neural 3D spatial coding?

Finally, we asked whether any of these spatiotemporal neural signatures of 3D visual processing might carry behavioral relevance in terms of 3D visual perception. Although our main task here involved passive viewing of 3D stimuli, each participant also participated in a behavioral session of the 3D cube adjustment task. We thus performed an exploratory analysis examining whether individual differences in depth perception were related to neural encoding of 3D-related information. For each participant, we extracted the slope parameter α (depth magnitude gain) from the Cube Adjustment Task, which reflects how strongly binocular disparity must be scaled to achieve a subjectively matched depth distance (perceived depth scaling). We then correlated these behavioral depth magnitude gains with each participant's maximum neural representational similarity (across fMRI searchlight units; EEG had too few subjects to be meaningful for this analysis) for different neural measures including the five depth- or 3D-related spatial features and the average of the four 3D features, the four different 3D coordinate systems, and 3D geometric distance and unique 3D geometric distance representations.

Among these measures, only the fMRI representation of the 3D radius feature (*r-3D*) showed a significant positive correlation with behavioral slope (r=.7406, p=.0143), indicating that participants with steeper disparity-to-depth mappings exhibited stronger neural encoding of 3D radius. All other correlations were non-significant (Figure S4), though we caution too much emphasis on this exploratory analysis of small sample size. That said, this finding suggests an intriguing potential link between individual differences in perceived depth magnitude and the strength of 3D spatial representations in the human brain.

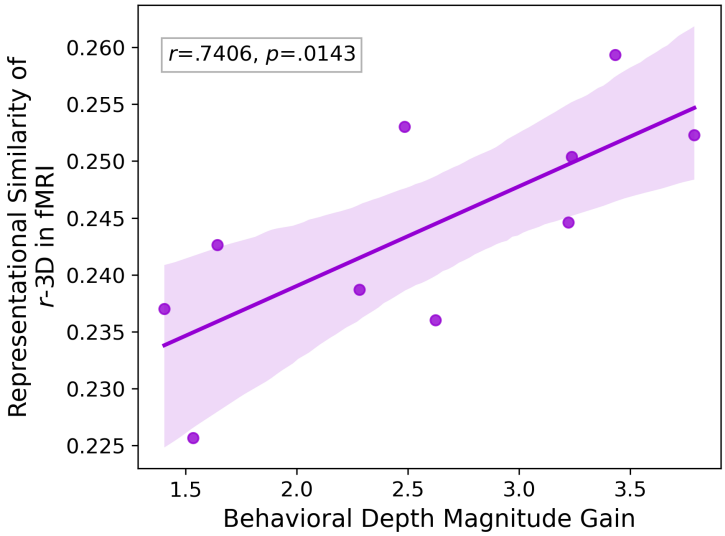

Figure 8 Correlation between behavioral depth magnitude gain and neural 3D radius encoding in fMRI.

Each dot represents one participant. Shaded area indicates 95% confidence interval for the fitted regression line.

Discussion

In this study, we combined high temporal-resolution EEG and high spatial-resolution fMRI recordings, computational approaches, and perceptual depth in binocular disparity to unfold how the human brain encodes 3D spatial location. Our novel findings translate into three principal contributions – methodological, theoretical, and open-science. Methodologically, we developed a multimodal framework combining individualized depth calibration, dense 3D spatial sampling, and partial correlation-based RSA applied to both EEG and fMRI. This allowed us to disentangle the neural representation of a rich set of spatial features across multiple 2D and 3D coordinate systems, and to identify feature-specific, coordinate-preferred, and geometric representations in time and brain space. Theoretically, our results demonstrate that the human brain flexibly encodes spatial information using multiple spatial features in multiple coordinate systems, including unique contributions of not only 2D features but also depth and 3D features, each expressed at different temporal stages and in distinct brain regions. We further provide evidence for higher-order spatial integration, suggesting that the human brain represents not only isolated feature dimensions but also relational geometric structure in 2D and 3D space. Together, our findings suggest that the human brain initially builds representations of 2D spatial features and 2D geometric distances, then 3D feature representations, and finally builds structured, unified representations of 3D spatial location (Figure 9). From an open-science perspective, we present the first publicly available large-scale neuroimaging dataset of 3D visual perception, which includes individualized depth calibration and multimodal (EEG and fMRI) recordings. We hope this resource will support future investigations into neural representations of spatial location and coordinate-system dynamics in real-world perception and behavior.

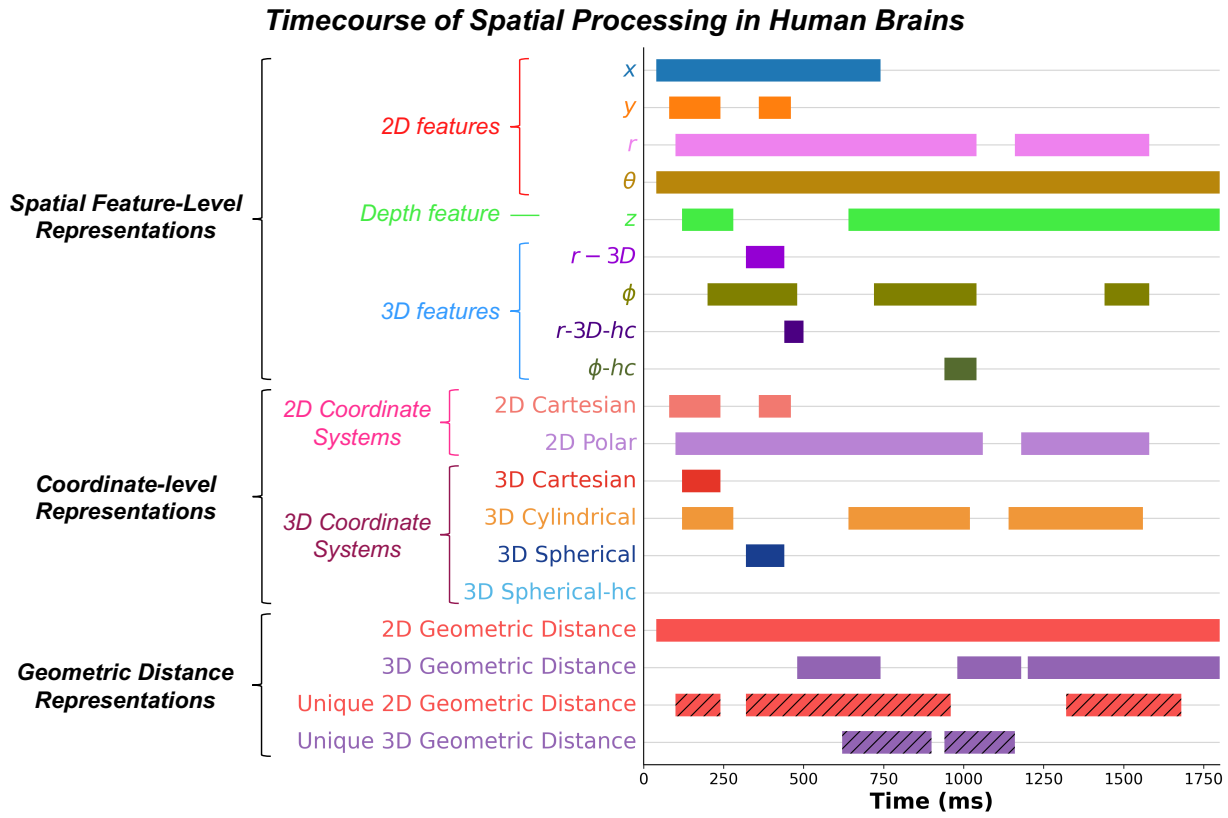

Figure 9 Time courses of spatial processing in human brains.
Summary schematic of the time courses of significant spatial feature-level representations, coordinate-level complete representations, and geometric distance representations (permutation test, cluster-corrected, $p < .01$).

Previous studies have primarily focused on a limited set of classic 2D or depth features (J. M. Carlson et al., 2011; Golomb & Kanwisher, 2012; Graumann et al., 2022; Kravitz et al., 2010). Most previous studies have typically analyzed these dimensions independently (Alvarez et al., 2021; Bridge & Parker, 2007; Finlayson et al., 2017), without quantitatively relating perceived depth distances to horizontal and vertical distances in 2D space. In contrast, our study systematically incorporates a broad set of spatial features defined across multiple coordinate systems in both 2D and 3D space – not only including 2D and depth features, but also extending to previously unexamined 3D features such as 3D radius and 3D polar angle. This was made possible by our ability to perceptually calibrate and precisely quantify the 3D positions of stimuli in 3D space. Critically, we employed a denser sampling of spatial locations (64 in total) within 3D space, which enabled us to disentangle overlapping feature representations and robustly analyze their contributions in human brains. This comprehensive feature space allowed us to consider multiple types of 3D geometric relationships rather than only comparing 2D and depth axes, allowing for a more complete assessment of how the brain encodes spatial information. Additionally, our use of partial correlation representational similarity analysis enabled us to

isolate the unique contributions of each spatial feature, going beyond standard correlation-based approaches that may conflate shared variance among features and misestimate the representational patterns (see correlation-based RSA results in Figure S5-S8). By including a richer and more diverse set of features, we were able to disentangle the unique neural representations of different spatial features.

Our results reveal a clear gradient in both time and brain space: 2D features (e.g., *x*, *y*, *r*, θ) emerged early in the EEG time course and were broadly represented across occipital cortex in fMRI. In contrast, the depth feature (*z*) appeared later in time and showed more restricted spatial activation, which overlapped with parietal and dorsal regions. 3D features (e.g., *r-3D*, Φ, *r-3D-hc*, and Φ*-hc*) showed anatomically distinct representations in fMRI and, although no significant time window when averaged together, individual 3D features exhibited distinct temporal profiles in EEG. From a temporal perspective, prior studies have shown that 2D position could be decoded from EEG signals within very early visual processing (Graumann et al., 2022; Hogendoorn & Burkitt, 2018), consistent with our findings of 2D features. However, to our knowledge, no previous EEG or MEG studies have examined the temporal dynamics of depth or 3D spatial features. From a spatial perspective, our fMRI results support and substantially extend prior findings that spatial encoding shifts from 2D to depth representations along the visual hierarchy (Finlayson et al., 2017; Henderson et al., 2019). Indeed we showed in a supplementary analysis a convincing replication of Finlayson et al., 2017's analysis (correlation-based MVPA instead of partial correlation-based RSA) and results (spatial transition from 2D (the average of x and y) to depth (z) along the visual hierarchy; Figure S1). Crucially, our study extends these findings in several important ways. First, we employed a much denser sampling of 64 3D spatial locations, which enabled us to characterize a richer set of spatial features across multiple coordinate systems – including not only traditional 2D and depth features, but also 3D features in multiple 3D coordinate systems. Second, we introduced integration analyses that revealed higher-order encoding of relational spatial geometry, beyond individual feature encoding. Third, by combining EEG and fMRI, we provided a multimodal spatiotemporal map of 3D spatial representation, revealing how different spatial dimensions emerge over time and are distributed across brain regions. Together, these advances allowed us to capture both the temporal dynamics and anatomical organization of spatial representations with greater specificity than prior studies.

A key finding from our study is that the human brain may not rely on a single coordinate system of spatial encoding. Instead, we found temporally and spatially distinct patterns for different coordinate systems, including 2D Cartesian, 2D Polar, 3D Cartesian, 3D Cylindrical, 3D Spherical, and 3D Spherical-hc coordinates. Our EEG results demonstrated that coordinate systems may differ in their representations over time, with a progressive timecourse across coordinate systems based on their complete representations, from early 2D Cartesian representations, to 2D Polar, followed by 3D Spherical, 3D Spherical-hc, and finally 3D Cylindrical representations. Notably, 2D Polar and 3D Cylindrical coordinate representations persisted longer in time. In parallel, our fMRI results showed that Cartesian- and Polar-format

spatial representations were expressed in distinct cortical regions. Although we identified only small patches of cortex showing a complete representation of 3D Cartesian – i.e., significant encoding of *x*, *y*, *z* within the same region – we did not observe complete representations of other 3D coordinate systems. This may reflect functional specialization, where different 3D features are processed across distributed neural populations, rather than being jointly encoded in a single region. Alternatively, it may be due to limited statistical power in detecting more spatially distributed or subtle effects, particularly for formats with higher redundancy or interdependence. Together, these findings suggest that the brain may maintain multiple spatial codes in parallel, which may suggest that our visual system may flexibly switch between coordinate systems or apply multiple systems simultaneously, depending on the computational demands of perception or action.

Beyond encoding individual features, a key novel contribution of our study is the identification of 2D and 3D geometric distance representations. Using integration RDMs based on 2D or 3D Euclidean distances and partial correlations, we found reliable evidence of 2D and 3D spatial integration in both EEG and fMRI data, which reflects the combined geometric relationships in 2D or 3D space rather than the individual feature dimensions. While 2D geometric distance was broadly represented, 3D geometric distance emerged only in specific regions, such as the parahippocampal cortex, and at middle and late timepoints in EEG. Interestingly, the parahippocampal cortex exhibited significant unique 3D geometric distance representation only after controlling for feature-level RDMs (Figure 7C), but not in the initial analysis (Figure 6D). This pattern likely reflects the removal of noisy variance shared with low-level spatial features, revealing a residual high-level 3D integration signal in this region. Such a dissociation is consistent with the parahippocampal cortex's proposed role in encoding abstract spatial relationships, such as real-world navigation and scene perception (Aminoff et al., 2013; R. Epstein et al., 1999, 2003; R. A. Epstein, 2008), beyond simple feature-based metrics. Importantly, our geometric distance RDMs were designed to capture the spatial relationship among stimuli in terms of their relative locations – how far each stimulus is from the others – rather than their absolute positions with respect to the reference point. This finding aligns with previous studies emphasizing the brain is sensitive not only to absolute location but also to relative spatial information (Hayworth et al., 2011; Roth, 2016), which underlies functions such as spatial comparison, grouping, and navigation (Berens et al., 2021; Byrne et al., 2007; Roth, 2016; Uchimura et al., 2015). Together, our results may suggest that human brains encode not only feature-based spatial information but also higher-order spatial structure that could capture relational geometry among multiple objects in space.

In addition, in an exploratory analysis we observed that individual differences in subjective depth perception from the behavioral Cube Adjustment Task were positively correlated with the neural encoding strength of fMRI representations of the 3D radial spatial feature (*r-3D)*. This finding suggests a potential link between perceptual scaling and neural coding of 3D radius (distance) information that participants who require larger disparity changes to perceive equivalent depth

distances may exhibit stronger neural coding of 3D radius information. While intriguing, due to the modest sample size, future studies with larger cohorts are needed to determine whether this relationship reflects a robust and generalizable neural signature of individual variability in depth perception.

Building on these findings, an important open question concerns the reference frames that underlie these spatial codes. Our analyses primarily defined locations relative to a gaze-centered origin, consistent with retinotopic or egocentric representations, with the exception of the head-centered Spherical (Spherical-hc) coordinates. However, because participants maintained central gaze fixation, our paradigm does not directly dissociate retinotopic from spatiotopic encoding, nor egocentric from allocentric coding (Burgess et al., 2007; Byrne et al., 2007; Crespi et al., 2011; Duhamel et al., 1997; Filimon, 2015; Gardner et al., 2008; Golomb & Kanwisher, 2012; Klatzky, 1998; Zaehle et al., 2007). Nonetheless, the involvement of regions such as the parahippocampal cortex—known for allocentric and scene-based representations (Aguirre et al., 1996; Ekstrom et al., 2014; Parslow et al., 2004; Rolls, 2020; Zaehle et al., 2007) – raises the possibility that multiple reference frames may coexist across cortical systems. Future work could explicitly manipulate fixation or use immersive environments to clarify how these reference frames interact in supporting 3D spatial perception.

Beyond the specific findings reported here, we believe the broader impact of this project lies in the resources and methodological framework it offers. To our knowledge, this study introduces the first large-scale, multimodal neuroimaging dataset for 3D spatial perception, and we provide a first- and novel-step in investigating the spatiotemporal neural representations of 3D visual perception in the human brain. We have publicly released this data, which we hope will serve as a foundation for future research efforts – not only to further advance our understanding of the neural mechanisms underlying 3D perception, but also to support translational applications such as decoding spatial location from brain activity in brain-computer interface (BCI) contexts across both cognitive neuroscience and bioengineering domains. Importantly, by making this dataset openly available, we also aim to enable the community to test alternative hypotheses and different models of 3D space representation, including approaches or feature dimensions beyond those we have explored here. In addition to the dataset, our work offers a generalizable methodological framework that integrates individual perceptual calibration, dense 3D spatial sampling, and multimodal RSA to probe high-resolution spatial encoding across both time and brain space.

While our study provides new insights into spatiotemporal encoding of 3D spatial information, several limitations should be acknowledged and addressed in future work. One potential concern in our study is that eccentricity-based cortical magnification might influence activation strength across the visual cortex (Deyoe et al., 1996; Dougherty et al., 2003; Harvey & Dumoulin, 2011; Sereno et al., 1995; Wandell et al., 2007). This issue arises because our stimulus locations were uniformly distributed in Cartesian coordinates, which do not account for the non-linear expansion of cortical surface area near the fovea. As a result, representational analyses relying on

differences in neural activation patterns across the visual field may be biased. To mitigate this issue, our RSA computations were based on Spearman partial correlations, a rank-based method that is robust to differences in scale and nonlinearity, and controls for shared variance across all feature models. This approach could minimize the influence of global activation magnitude and cortical surface area differences. Nevertheless, the cortical magnification may still lead to underestimation or distortion of representational similarity results in certain regions. Future experiments could consider eccentricity-informed spatial sampling to complement our findings. Another limitation of the present study is that our exploration of 3D spatial encoding was restricted to depth defined by binocular disparity. While binocular disparity is a powerful and well-studied cue to depth, real-world 3D perception relies on the integration of multiple depth cues, including motion parallax, texture gradients, relative size, shading, occlusion and perspective (Howard, 2012; Welchman, 2016). It remains unknown whether the spatiotemporal dynamics we observed for disparity-defined 3D representations generalize to other depth cues or to cue-integrated depth perception. Future work should therefore examine how different depth cues are represented individually and jointly in the brain, and how their integration shapes holistic 3D spatial representations. Also, an important direction for future research concerns how task demands modulate spatial encoding. In the current study, we employed a passive viewing paradigm to establish a baseline map of 3D spatial representations under minimal cognitive engagement. However, future studies could incorporate multiple active tasks to test how spatial representations are dynamically shaped by behavioral goals and cognitive context. Future work can build on our methodological framework to investigate coordinate-system dynamics in active viewing, naturalistic environments, or memory-guided navigation. These extensions will help us clarify not only why human brains apply different systems to represent spatial locations, but also how 3D spatial representations interface with eye movements, motor control, attention, and decision-making processes.

Methods

Participants

Ten participants participated in the study (7 females and 3 males, mean age = 26.0 ± 5.5 years) for monetary compensation ($15/hr for the behavioral session, and $20/hr for the fMRI and EEG sessions). Eight participants each completed one behavioral session, two EEG sessions, and two fMRI sessions in five separate days. The other two participants completed one behavioral session and two fMRI sessions in three separate days. All participants first completed the behavioral session, and subsequently did the neuroimaging sessions, scheduled according to participant availability and the scanner and EEG lab calendars. All participants had normal or corrected-

normal vision, and were prescreened for MRI eligibility. The study protocol was approved by The Ohio State University Biomedical Sciences Institutional Review Board.

Our sample size was determined based on prior neuroimaging studies using large-scale EEG and/or fMRI to investigate high-dimensional neural encoding (Allen et al., 2022; Chang et al., 2019; Gifford et al., 2022; Hebart et al., 2023). Given the within-subject design and the large number of trials per participant across five sessions, the dataset affords high sensitivity to spatiotemporal representational structure at both individual and group levels. While a larger sample may improve generalizability, our findings reflect robust effects across participants and provide a strong foundation for future replication and extension studies.

General Experimental Setup

Dynamic random dot stereogram stimuli (RDS) were generated using Psychtoolbox extension (Brainard, 1997) for MATLAB (Math Works). Depth from binocular disparity was achieved using red/green anaglyph glasses paired with Psychtoolbox's stereomode. For all the experiments in our study, we created a 3D space composed of a low-contrast, large background field (12° × 12°) placed at the central depth plane of the screen, framed with additional depth cues making a 3D reference frame. The frontmost and backmost depth frames were rendered with disparities of +20 arcmin and -20 arcmin, respectively, creating a 40 arcmin total depth range. The background field consisted of static random dot stimuli (RDS; 5 dots/deg^2) comprised of light and dark gray dots (each sized 0.18° × 0.18°) on a mid-gray background. To enhance the 3D percept and provide a stable spatial reference across depth, we also displayed visual depth cues composed of vertical and horizontal grid lines framing the 3D space outside the stimulus area. This grid formed a perspective-like scaffold, giving observers a consistent visual structure suggestive of depth to encourage perception of a 3D space. The main experimental stimuli were cubes of high-contrast dynamic RDS (each cube sized 2.4° × 2.4° × 2.4°, composed of 0.18° × 0.18° white and back grey dots, with 10 dots/deg^2), presented at different locations within this 3D space, as described in the sections below.

For behavioral and EEG sessions, stimuli were presented on a 21-inch LCD monitor (resolution 1920 × 1080 at 240 Hz), and participants were seated at a chinrest 74 cm from the monitor. For fMRI sessions, stimuli displayed with a DLP projector onto a screen mounted in the rear of the scanner (resolution 1280×1024 at 60 Hz), and participants viewed from a distance of 74 cm via a mirror above their heads attached to the head coil.

Behavioral Session

The behavioral session included three tasks. The first two tasks were used to confirm that the participants could accurately perceive and discriminate different depth distances with RDS from binocular disparity, and the third task was used to measure different disparity parameters corresponding to different depth distances. In addition, the extensive exposure during the whole

behavioral session served to acclimate participants to the depth information in these displays, enhancing their sensitivity to depth cues in the subsequent main tasks.

In Task 1 – Single stimulus depth judgement task, participants viewed a single 2.4° × 2.4° square patch of dynamic RDS presented at the 2D center of the screen, appear either in front or behind the screen (fixation) plane. The stimulus depth was set to one of four disparity levels: +15, +5, -5, -15 arcmin (relative to the fixation depth plane). Participants were asked to judge whether the stimulus appeared in front of (closer to them) or behind (further away) the screen. Each trial was self-paced with no time limit and no eye movement restrictions. After participants responded, feedback indicating “correct” or “incorrect” was displayed, and the trial ended. Participant completed four blocks in total, each consisting of 24 trials (4 depth levels × 6 repetitions), with trials presented in a fully randomized order. After completing the first two blocks, participants reversed the orientation of their glasses (i.e., left-red/right-green or left-green/right-red). In one block per orientation, the key mapping was “A” for “in front” and “L” for “behind”, while in the other, the mapping was reversed (“L” for “in front”, “A” for “behind”).

In Task 2 – Two-stimulus depth judgment, participants viewed two dynamic RDS patches presented simultaneously to the left and right of central fixation (each 2.4° × 2.4°, with center positions 3° horizontally from fixation). The two RDS patches were presented at different depth levels, forming one of six predefined disparity pairs: left: −15 arcmin / right: −5 arcmin; left: −5 arcmin / right: −15 arcmin; left: −5 arcmin / right: +5 arcmin; left: +5 arcmin / right: −5 arcmin; left: +5 arcmin / right: +15 arcmin; left: +15 arcmin / right: +5 arcmin. The two stimuli could both appear in front of, behind, or straddling the fixation (screen) depth plane. Participants were instructed to judge the relative depth of the two stimuli. At the beginning of each block, participants were informed whether they should report which stimulus was closer or which was farther. For example, in a "closer" block, participants pressed the 'A' key if they thought the left stimulus was closer, or the 'L' key if they thought the right stimulus was closer. Each trial was self-paced with no time limit and no restriction on eye movements. Once the participant responded, feedback (“correct” or “incorrect”) was displayed, and the trial ended. Participants completed four blocks, each consisting of 12 trials (6 depth pairs × 2 repetitions), presented in randomized order. After two blocks, the color orientation of the anaglyph glasses was reversed (left-red/right-green or left-green/right-red; counterbalanced across participants). For each anaglyph orientation, one block required judging which stimulus was closer, and the other required judging which was farther.

All our ten participants showed > 90% accuracy in both Task 1 and 2, which verified that they were sensitive to the depth triggered by binocular disparity. Then they did the Task 3.

In Task 3 – Cube adjustment task, participants viewed four RDS patches (each 2.4° × 2.4°) presented simultaneously in the upper-left, upper-right, lower-left, and lower-right quadrants of the screen. These four stimuli corresponded to the four vertices of a square shape defined by their farthest edges from fixation. Across trials, the horizontal and vertical distances of this “virtual”

square were simultaneously manipulated to form one of 11 physical sizes: 5.4°×5.4°, 6.0°×6.0°, 6.6°×6.6°, 7.2°×7.2°, 7.8°×7.8°, 8.4°×8.4°, 9.0°×9.0°, 9.6°×9.6°, 10.2°×10.2°, 10.8°×10.8°, or 11.4°×11.4°. Crucially, the four stimuli were arranged such that two diagonally opposite patches were presented in front of the fixation plane and the other two behind the fixation plane, with equal absolute disparities (i.e., symmetrical distance from fixation). The initial disparity was randomly selected on each trial. Participants performed a cube adjustment task: using the up/down arrow keys, they simultaneously moved the front and back planes either closer to or farther from the fixation plane, until they perceived the four vertices as forming a cube in 3D space. Note that while depth positions varied, the 2D square size on the screen remained fixed within each trial. Each trial was self-paced with no time limit and no restriction on eye movements. Participants pressed the "Enter" key to confirm when they believed the configuration matched a cube, thus ending the trial. Participants completed two blocks of 66 trials each (11 horizontal/vertical square sizes × 2 depth arrangements (half with top-left and bottom-right stimuli in front, and half with them behind) × 3 repetitions), with trials presented in random order. After the first block, the orientation of the anaglyph glasses was reversed (left-red/right-green or left-green/right-red). This task allowed us to determine the individualized perceptually-matched binocular disparities corresponding to a range of reference distances from the fixation plane, ranging from 2.7° to 5.7° in 0.3° increments (i.e., half of the cube's front-back extent). These distances reflect the amount of disparity required for a specific participant to perceive the four stimuli as forming a cube in 3D space, providing a participant-specific calibration of depth perception in stereoscopic space, allowing us to match the depth and 2D distances sampled in the main tasks.

Individualized Depth Distance Quantification

To establish a participant-specific mapping between perceived depth distance and binocular disparity, we analyzed data from the Cube Adjustment Task (Task 3). We fit a linear function to each participant's data using least-squares regression, with perceptual matched disparity as the dependent variable and reference distance as the independent variable. Given that a depth distance of 0° should yield zero disparity, we constrained the fit to pass through the origin, yielding a function of the form $Y_{\text{disparity}} = \alpha X_{\text{reference\_distance}}$. The slope α captures the participant-specific transformation from reference distance (in degrees of visual angle) to perceptually-matched binocular disparity (in arcmin). This slope can be interpreted as an individualized "depth magnitude gain", indexing how strongly binocular disparity must be scaled to achieve a subjectively matched depth distance (perceptual depth scaling). Figure S9 shows the model fitting results for all ten subjects.

This individualized linear model was used to compute the disparity corresponding to any desired depth distance for each participant, enabling precise, perceptually calibrated depth manipulation across all subsequent experimental sessions.

EEG Session

In each EEG session, during each trial, participants viewed a small dynamic RDS cube stimulus (2.4° × 2.4° × 2.4°) presented at one of 64 possible 3D spatial locations. The 64 locations were defined by a 4 x 4 x 4 (horizontal x vertical x depth) grid with fixation at the origin ([0° 0° 0°]). The 64 possible stimulus locations were the intersections at the following grid coordinates: horizontal *x* position: -4.5°, -1.5°, 1.5°, 4.5° × vertical *y* position: -4.5°, -1.5°, 1.5°, 4.5° × depth *z* position: -4.5°, -1.5°, 1.5°, 4.5°. The exact amount of binocular disparity required to produce each depth position was individually calibrated for each participant using the linear function derived from the Cube Adjustment Task (behavioral Task 3). The stimulus was a cube centered on these locations, and our design of the cube size and spacing was to ensure that stimuli at adjacent locations were spatially independent, with no overlap in 3D space. Such separation was critical for disentangling neural activity patterns associated with individual spatial locations.

Each stimulus was presented for 1.5 s, followed by a variable interstimulus interval (ISI) of 0.5-0.7 s (uniformly jittered). Participants were instructed to maintain central fixation at all times. The fixation cue consisted of two concentric dark gray circles (outer diameter: 0.45°; inner diameter: 0.15°) with a light gray crosshair at the center, presented at coordinates [0,0,0]. To ensure attention and maintain fixation, participants performed a fixation-change detection task: on a small proportion of trials, the central cross briefly changed into an “X” shape, and participants were instructed to press the spacebar upon detecting the change. Participants completed 18 blocks in each session, each consisting of 128 trials (64 spatial locations × 2 repetitions), with trials presented in fully randomized order. To control for any low-level color or contrast differences between eyes or monocular-based cues, participants reversed the orientation of the red/green anaglyph glasses across sessions, such that the red filter was on the left eye for once session and the right eye for the other. The order of initial color assignment was counterbalanced across participants. This flip manipulation counterbalanced which eye received the red versus green images across sessions, eliminating simple monocular or color-based cues. Importantly, near and far positions had equal disparity magnitudes. Thus, our results cannot be explained by absolute disparity magnitude alone, but instead reflect neural coding of the perceived spatial location.

fMRI Session

In each fMRI session, participants completed ten to twelve main task runs (6.1 mins per run). Similar to the EEG blocks, each run consisted of 128 trials (64 spatial locations × 2 repetitions). On each trial, a dynamic RDS cube stimulus (2.4° × 2.4° × 2.4°) was presented for 1.5 s at one of the 64 spatial locations (same exact stimulus positions and presentation time as EEG task). Interleaved with the 128 stimulus trials were 42 blank trials (no stimulus presented), inserted randomly such that no more than two consecutive blank trials occurred, to achieve sufficient jitter for modeling rapid event-related fMRI BOLD activity. Participants were instructed to maintain fixation at the center of the screen and perform a fixation change detection task,

pressing the spacebar whenever the central fixation symbol changed from a "+" to an "x"; these fixation changes could occur on blank trials as well as stimulus trials.

In addition to the main task runs, participants also completed two retinotopic mapping runs using standard rotating wedge stimuli (Engel et al., 1994; Sereno et al., 1995). High-contrast radial checkerboard patterns were presented as 60° wedges and flickered at 4 Hz. Maximal eccentricity was 16° and the central 1.6° foveal region was not stimulated (except for a central fixation point). One run rotated clockwise, and the another run rotated counter-clockwise through 7 cycles with a period of 24 s/cycle. During these runs, participants fixated at the center of the display and presses a button every time the black fixation dot dimmed to gray.

As with the EEG sessions, the anaglyph glasses direction was reversed across fMRI sessions, with the order randomized across participants.

Eye Tracking

Eye position was monitored with an EyeLink 1000 eye-tracking system in both EEG lab and the fMRI scanner. The eye tracker was calibrated using a nine-point grid method at the beginning of each neuroimaging session and re-calibrated as necessary. For fMRI, due to occasional loss of pupil signal in the scanner environment, fixation was primarily monitored in real time via the eye-tracking display. Trials were not aborted or removed for poor fixation, but we visually monitored and verified that participants were maintaining central fixation on the vast majority of trials. For EEG, if gaze deviated by more than ~2° from fixation during the 1.5 s stimulus presentation, the trial was immediately aborted and a new trial started.

EEG Acquisition

EEG experiments were carried out at the EEG lab in Department of Psychology at The Ohio State University. EEG data were recorded using an elastic cap (Brain Products ActiCap) with 64 active electrodes (including one online reference channel: FCz, and other 63 channels: FP1, FP2, AFz, AF3, AF4, AF7, AF8, Fz, F1, F2, F3, F4, F5, F6, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FT7, FT8, FT9, FT10, Cz, C1, C2, C3, C4, C5, C6, T7, T8, CPz, CP1, CP2, CP3, CP4, CP5, CP6, TP7, TP8, TP9, TP10, Pz, P1, P2, P3, P4, P5, P6, P7, P8, POz, PO3, PO4, PO7, PO8, Oz, O1, O2) arranged in the standard 10-20 layout, and a BrainVision actiCHamp amplifier at a sampling rate of 1000 Hz with the online filtering (between 0.1 Hz and 100 Hz). Electrode impedances were reduced to <20 kΩ before the commencement of each experiment session.

fMRI Acquisition

fMRI experiments were carried out at The Ohio State University Center for Cognitive and Behavioral Brain Imaging with a Siemens Prisma 3T MRI scanner using a 32-channel phase array receiver head oil. Functional data were acquired using a T2-weighted gradient-echo planar imaging (EPI) sequence (TR = 2000 ms, TE = 30 ms, flip angle = 72°; 2 × 2 × 2 mm voxel size; 72 axial slices; no gap). The acquisition was aligned to the anterior commissure-posterior

commissure (AC-PC) plane. Multiband acceleration was applied using the CMRR mbep2d_bold sequence with a multiband factor of 3. A high-resolution T2-weighted turbo spin echo (TSE) sequence was acquired for hippocampal subfield segmentation (TR = 4800 ms, TE = 106 ms, flip angle = 135°; voxel size = 0.5 × 0.5 × 2 mm). Also, we collected a high-resolution MPRAGE anatomical scan (1 mm^3) for each participant.

EEG Preprocessing

Offline EEG preprocessing was conducted in Python, using the MNE-Python package (Gramfort et al., 2013) along with customized scripts based on a previously published pipeline (Lu et al., 2024). EEG data from two recording sessions for each participant were merged into a single dataset. A band-pass filter from 0.1 to 30 Hz was applied, and independent component analysis (ICA) was used to identify and remove artifacts related to eye blinks and eye movements (Drisdelle et al., 2017; Jung et al., 2000). Signals from channels TP9 and TP10, placed over the left and right mastoids, were used for re-referencing. The continuous EEG data were segmented into epochs from -200 ms to 1800 ms relative to stimulus onset. Baseline correction was performed by substracting the mean voltage of a 100 ms pre-stimulus period (-100 to 0 ms) for each trial and channel. This resulted in a matrix of preprocessed EEG data for each participant with dimensions 4,608 trials × 63 channels × 2,000 timepoints.

To obtain location-specific event-related potentials (ERPs), we averaged the 72 repeated trials corresponding to each of the 64 stimulus location labels. The resulting data were then downsampled to 50 Hz by averaging every 20 consecutive timepoints. This procedure yielded an ERP matrix with dimensions 64 stimulus locations × 63 electrode channels × 100 timepoints for each participant, which was used for all subsequent analyses.

fMRI Preprocessing

fMRI data were preprocessed using Brain Vogager QX (Brain Innovation). Preprocessing included slice timing correction, head motion correction, temporal filtering, and normalization to Talairach space (Talairach & Tournoux, 1988). No spatial smoothing was applied to the data used for representational similarity analysis. A whole-brain random-effects general linear model (GLM) was run for all main task runs across fMRI sessions to calculate beta weights for each voxel, for each spatial location condition, for each participant. This yielded a beta-weight matrix for each participant with dimensions 64 stimulus locations × 432,216 voxels. Each participant's cortical surface for each hemisphere was inflated and flattened into cortical surface space for retinotopic mapping.

Retinotopic regions of interest (ROIs) including V1v, V1d, V2v, V2d, V3v, V3d, and V4 were functionally defined based on individual retinotopic mapping data. Additional ROIs—V3a, V3b, IPS0, IPS1–5, VO1, VO2, LO1, LO2, MST, MT, SPL1, and FEF—were defined using maximum probability maps from the Wang et al. probabilistic atlas (Wang et al., 2014). Medial temporal lobe ROIs, including the parahippocampal cortex, hippocampus, and entorhinal cortex (ERC),

were anatomically defined using individual subject segmentations provided by FreeSurfer (Fischl, 2012).

Representational Similarity Analysis

Our key analyses were based on representational similarity analysis (Kriegeskorte et al., 2008). The subsections below explain the details of how we computed representational dissimilarity matrices (RDMs) based on hypothetical representational spaces and actual neural signals (fMRI or EEG), and how we measured the unique representation of each spatial feature using the partial correlation approach (Dobs et al., 2019; Lu & Golomb, 2024). All analyses were implemented using customized code adapted from the NeuroRA toolbox (Lu & Ku, 2020).

Hypothesis-based Representational Dissimilarity Matrices (RDMs)

We constructed nine hypothesis-based spatial feature RDMs, each reflecting a distinct spatial feature dimension (Figure 1E).

For the *x*, *y*, and *z* RDMs, we extracted the respective Cartesian coordinates (horizontal, vertical, and depth positions) of the 64 stimulus locations along the X-, Y-, and Z-axes. Each RDM was constructed as a 64 × 64 matrix where each cell reflects the dissimilarity along the given spatial dimension between a pair of two stimulus positions. For example, to compute the spatial dissimilarity between a stimulus centered at position #18 (x=-1.5°, y=-4.5°, z=-1.5°) and a stimulus centered at position #51 (x=4.5°, y=-4.5°, z=1.5°), we would calculate the absolute differences in their x, y, and z locations, respectively. This would produce a dissimilarity value of 6 for that cell of the X-RDM, a 0 for the Y-RDM, and a 3 for the Z-RDM. These hypothesis-based RDMs reflect the representational similarity patterns expected if the brain represented purely horizontal (x), vertical (y), or depth (z) information.

For the *r* and θ RDMs, we analogously extracted the respective Polar coordinates (radius and polar angle) of the 64 stimulus locations along the r- and θ-axes. For *r,* we computed the 2D radial distance from each location to the fixation point on the screen. For θ, we calculated the corresponding polar angle relative to the positive X-axis within the 2D plane. For each of these, the RDMs were computed as the absolute difference in radius or polar angle between each pair of stimulus conditions.

For the *r-3D* RDM, we computed the 3D radial distance from each stimulus location to the fixation point on the screen in the 3D space. For the Φ RDM, we calculated the angle between the 3D direction vector from the fixation point on the screen to each location and the positive Z-axis. For the *r-3D-head-centered* RDM and the *Φ-head-centered* RDM, we computed the 3D Euclidean distance (3D head-centered radius) and 3D head-centered polar angle from each location to the participant's head position as the reference point in the 3D space. For each of these, dissimilarity values in the RDM were defined as the absolute differences in these dimensions between each pair of conditions.

Additionally, we constructed two additional hypothesis-based geometric distance RDMs, each reflecting the combined geometric relationships among different spatial locations in 2D or 3D space instead of the representations of individual spatial feature dimensions. For 2D geometric distance RDM, we computed the 2D Euclidean distance in the 2D space between every pair of stimulus locations as the dissimilarity. Similarly, for 3D geometric distance RDM, we computed the 3D Euclidean distance in the 3D space between every pair of stimulus locations as the dissimilarity.

Neural RDMs

EEG electrodes and fMRI voxels each aggregate signals from a mixture of neural sources and spatial locations, making it important to examine both univariate and multivariate dissimilarity measures of representational dissimilarity to consider two types of potential representational formats in the neural recordings. Univariate measures, such as mean amplitude differences, are sensitive to overall or large-scale differences in response strength, while multivariate measures, such as voxel-wise correlation, capture distributed spatial patterns across the scalp (for EEG) or a region of interest (from fMRI). Accordingly, we computed two types of RDMs for each timepoint in the EEG time series and each searchlight unit in the fMRI volume: one based on amplitude differences – the amplitude-based RDM, and one based on correlation distance – the pattern-based RDM.

EEG timepoint-by-timepoint RDMs: For the amplitude-based RDMs, we discarded all spatial pattern information. At each timepoint, we computed the mean amplitude across all 63 channels for each stimulus location condition. The dissimilarity between any pair of conditions was then defined as the absolute difference between their mean amplitudes, yielding a 64 × 64 EEG amplitude-based RDM. For the pattern-based RDMs, we discarded absolute activation information by first z-scoring the values at each channel separately across all conditions. At each timepoint, we treated the 63-channel activation vector as a multivariate pattern and computed one minus the Pearson correlation coefficient between pairs of condition-specific vectors, resulting in a 64 × 64 EEG pattern-based RDM.

fMRI searchlight RDMs: For the amplitude-based RDMs, we discarded spatial pattern information within each searchlight unit. For each stimulus location condition, we computed the mean activation (beta value) across all 27 voxels within a given searchlight 3 × 3 × 3 cube unit. The dissimilarity between any pair of conditions was defined as the absolute difference between their mean activations, resulting in a 64 × 64 amplitude-based RDM for each searchlight. For the pattern-based RDMs, we discarded absolute activation information by first z-scoring voxel responses across conditions within each searchlight. Each condition-specific response was treated as a multivoxel pattern, and dissimilarity between pairs of conditions was computed as one minus the Pearson correlation between their voxel-wise activation vectors, yielding a 64 × 64 pattern-based RDM per searchlight unit (across brain space).

fMRI ROI-based RDMs: In addition to the searchlight RDMs, we computed a set of RDMs within predefined ROIs. For each ROI and participant, we extracted the voxel-wise beta weights for all 64 spatial location conditions. For the amplitude-based RDMs, we discarded voxel-level spatial pattern information by computing the mean beta value across all voxels in the ROI for each condition. Dissimilarity between condition pairs was defined as the absolute difference between their mean activation values, resulting in a 64 × 64 amplitude-based RDM per ROI. For the pattern-based RDMs, we first z-scored the voxel-wise beta values across conditions within each ROI to remove mean-level activation differences. Each condition-specific voxel activation pattern was then compared using one minus the Pearson correlation coefficient, yielding a 64 × 64 pattern-based RDM for each ROI.

Partial correlations between neural and feature RDMs

To evaluate how human brains represent different spatial features across time (EEG) and brain space (fMRI), we calculated the representational similarity between neural RDMs – temporal EEG RDMs, fMRI searchlight RDMs, or fMRI ROI RDMs – and the nine hypothesis-based feature RDMs described above. To isolate the unique contribution of each spatial feature and mitigate collinearity among feature RDMs, we used rank-based Spearman partial correlation analysis (Dobs et al., 2019; Lu & Golomb, 2024). Spearman correlation was chosen because it is robust to non-linear relationships and differences in scale across RDMs, making it well-suited for comparing representational dissimilarity structures. Specifically, for each 64 × 64 RDM, we extracted the upper triangular values excluding the diagonal (2,016 dissimilarity values) and reshaped them into a 1×2,016 vector. We then computed the partial correlation between a given neural RDM vector and each target feature RDM vector (e.g., the *x* RDM), while statistically controlling for the remaining eight vectors corresponding to the other eight feature RDMs. This procedure removed shared variance with other spatial features and yields a measure of the unique representational similarity for the target feature dimension for each EEG timepoint and fMRI searchlight or ROI.

Similarly, to evaluate whether there is also evidence of processing 2D and 3D location holistically which reflects the combined geometric relationships in 2D or 3D space rather than the individual feature dimensions, we calculated the partial correlation between a given neural RDM vector and each geometric distance RDM vector (2D or 3D geometric distance RDM), while statistically controlling for the alternative geometric distance RDM. Additionally, we also conducted a version of this analysis taking the partial correlation between a given neural RDM vector and each geometric distance RDM vector, while statistically controlling for all feature RDM vectors as well. This procedure removed shared variance with all individual spatial features and the alternative integration representation and yields a measure of the unique representational similarity for the 2D or 3D geometric distance representation regardless of the feature- or coordinate-level encoding format.

As noted above, we constructed two versions of the neural RDMs at each EEG timepoint and fMRI searchlight/ROI – an amplitude-based RDM and a correlation-based RDM. Rather than choose one over the other, since they both capture potentially relevant aspects of the neural signal, we conducted the partial correlation analysis described above for both types of neural RDMs separately and then took the maximum. (Figures S10-S16 show the separate results from amplitude- and pattern-based RDMs.) We defined our robust summary estimate of the final representational similarity for each feature × timepoint (and feature × brain region) as the maximum of these two measures rather than the average because the two measures capture complementary aspects of the neural signal. By taking the maximum, we ensure that if one metric robustly reflects the underlying representation in a given time window or a given brain region – despite potential noise or variations in sensitivity – the overall analysis will not underestimate the representational strength. This conservative approach minimizes the risk of overlooking a genuine representation that may be captured preferentially by one metric over the other.

Statistical Significance: Permutation Tests

To assess the statistical significance of group-level representational similarity for each spatial feature, we performed permutation-based tests. For each participant and each neural RDM (EEG temporal RDMs, fMRI searchlight RDMs, or fMRI ROI RDMs), we first extracted the 2,016 unique dissimilarity values from the upper triangle of the 64 × 64 matrix and permuted these values 1,000 times to generate null RDMs. For each permutation, we computed the partial correlation between the permuted neural RDM and each of the nine model RDMs, while controlling for the remaining eight as described above. The resulting partial correlations were then averaged across participants to form a group-level null distribution for each spatial feature. To evaluate significance, we compared the true group-level partial correlation (based on unshuffled neural RDMs) to its corresponding null distribution. A feature was considered significantly represented if the true group-level similarity exceeded the 99th percentile of the null distribution (one-sided test, $p < 0.01$). For EEG temporal and fMRI searchlight analyses, we further applied cluster-based correction across contiguous timepoints or spatial units to control for multiple comparisons. For ROI-based analyses, no cluster correction was applied due to the absence of spatial or temporal continuity.

Data Availability

All data and code will be freely available post-publication on GitHub and OSF.

Acknowledgements

This work was supported by research grants from the National Institutes of Health (R01-EY025648).

References

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The Parahippocampus Subserves Topographical Learning in Man. *Cerebral Cortex*, *6*(6), 823–829.

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126.

Alvarez, I., Hurley, S. A., Parker, A. J., & Bridge, H. (2021). Human primary visual cortex shows larger population receptive fields for binocular disparity-defined stimuli. *Brain Structure and Function*, *226*(9), 2819–2838.

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*(8), 379–390.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology*, *86*(4), 2054–2068.

Berens, S. C., Joensen, B. H., & Horner, A. J. (2021). Tracking the Emergence of Location-based Spatial Representations in Human Scene-Selective Cortex. *Journal of Cognitive Neuroscience*, *33*(3), 445–462.

Bridge, H., Ip, I. B., & Parker, A. J. (2023). Investigating the human binocular visual system using multi-modal magnetic resonance imaging. *Perception*, *52*(7), 441–458.

Bridge, H., & Parker, A. J. (2007). Topographical representation of binocular depth in the human visual cortex using fMRI. *Journal of Vision*, *7*(14), 1–14.

Burgess, N., Barry, C., & O'Keefe, J. (2007). An oscillatory interference model of grid cell firing. *Hippocampus*, *17*(9), 801–812.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375.

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *NeuroImage*, *57*(4), 1608–1616.

Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. J. (2011). Spatial coding and

invariance in object-selective cortex. *Cortex*, *47*(1), 14–22.

Carvalho, J., Invernizzi, A., Ahmadi, K., Hoffmann, M. B., Renken, R. J., & Cornelissen, F. W. (2020). Micro-probing enables fine-grained mapping of neuronal populations using fMRI. *NeuroImage*, *209*, 116423.

Chang, N., Pyles, J. A., Marcus, A., Gupta, A., Tarr, M. J., & Aminoff, E. M. (2019). BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Scientific Data*, *6*(1), 1–18.

Chen, N., Chen, Z., & Fang, F. (2020). Functional specialization in human dorsal pathway for stereoscopic depth processing. *Experimental Brain Research*, *238*(11), 2581–2588.

Conner, I. P., Sharma, S., Lemieux, S. K., & Mendola, J. D. (2004). Retinotopic organization in children measured with fMRI. *Journal of Vision*, *4*(6), 10–10.

Crespi, S., Biagi, L., d'Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic Coding of BOLD Signal in Human Visual Cortex Depends on Spatial Attention. *PLOS ONE*, *6*(7), e21661.

Deyoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., & Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(6), 2382–2386.

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1), 1–10.

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision*, *3*(10), 1–1.

Drisdelle, B. L., Aubin, S., & Jolicoeur, P. (2017). Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory: ICA correction versus epoch rejection. *Psychophysiology*, *54*(1), 83–99.

Duhamel, J. R., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, *389*(6653), 845–848.

Ekstrom, A. D., Arnold, A. E. G. F., & Iaria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: Toward a network-based perspective. *Frontiers in Human Neuroscience*, *8*, 113369.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). FMRI of human visual cortex. *Nature*, *369*(6481), 525.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*(5), 865–876.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area:

Recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Filimon, F. (2015). Are all spatial reference frames egocentric? Reinterpreting evidence for allocentric, object-centered, or world-centered reference frames. *Frontiers in Human Neuroscience*, *9*, 160193.

Finlayson, N. J., Zhang, X., & Golomb, J. D. (2017). Differential patterns of 2D location versus depth decoding along the visual hierarchy. *NeuroImage*, *147*, 507–516.

Fischer, J., Spotswood, N., & Whitney, D. (2011). The Emergence of Perceived Position in the Visual System. *Journal of Cognitive Neuroscience*, *23*(1), 119–136.

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of Visual Space in Human Occipital Cortex Are Retinotopic, Not Spatiotopic. *Journal of Neuroscience*, *28*(15), 3988–3999.

Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, *264*, 119754.

Golomb, J. D. (2018). Representations of 3D visual space in human cortex: Population receptive field models of position-in-depth. *Cognitive Computational Neuroscience*.

Golomb, J. D., & Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cerebral Cortex*, *22*(12), 2794–2810.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 70133.

Graumann, M., Ciuffi, C., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). The spatiotemporal neural dynamics of object location representations in the human brain. *Nature Human Behaviour*, *6*(6), 796–811.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*(Volume 27, 2004), 649–677.

Grootswagers, T., Zhou, I., Robinson, A. K., Hebart, M. N., & Carlson, T. A. (2022). Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data*, *9*(1), 1–7.

Harvey, B. M., & Dumoulin, S. O. (2011). The Relationship between Cortical Magnification Factor and Population Receptive Field Size in Human Visual Cortex: Constancies in Cortical Architecture. *Journal of Neuroscience*, *31*(38), 13604–13612.

Hayworth, K. J., Lescroart, M. D., & Biederman, I. (2011). Neural Encoding of Relative Position. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1032–1050.

Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., Corriveau, A., Vaziri-Pashkam, M., & Baker, C. I. (2023). THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior.

ELife, *12*, e82580.

Henderson, M., Vo, V., Chunharas, C., Sprague, T., & Serences, J. (2019). Multivariate Analysis of BOLD Activation Patterns Recovers Graded Depth Representations in Human Visual and Parietal Cortex. *ENeuro*, *6*(4).

Hogendoorn, H., & Burkitt, A. N. (2018). Predictive coding of visual object position ahead of moving objects revealed by time-resolved EEG decoding. *NeuroImage*, *171*, 55–61.

Howard, I. (2012). *Perceiving in depth, volume 1: basic mechanisms*. Oxford University Press. https://books.google.com/books?hl=en&lr=&id=A26JAgAAQBAJ&oi=fnd&pg=PP1&dq=Howard,+I.P.,+2012.+Perceiving+in+DepthVolume+1+Basic+Mechanisms.+Oxford+University+Press.&ots=Ayj1XR31FY&sig=YtZJ-wOI3J-KAtZCLuVzxvM45-k

Ip, I. B., Minini, L., Dow, J., Parker, A. J., & Bridge, H. (2014). Responses to interocular disparity correlation in the human cerebral cortex. *Ophthalmic and Physiological Optics*, *34*(2), 186–198.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745–1758.

Klatzky, R. L. (1998). Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections. In *Spatial Cognition* (Vol. 1404, pp. 1–17). Springer, Berlin, Heidelberg.

Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-Level Visual Object Representations Are Constrained by Position. *Cerebral Cortex*, *20*(12), 2916–2925.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *4*.

Lu, Z., & Golomb, J. D. (2024). Human EEG and artificial neural networks reveal disentangled representations of object real-world size in natural images. *ELife*, *13*, RP98117.

Lu, Z., & Ku, Y. (2020). NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. *Frontiers in Neuroinformatics*, *14*, 61.

Lu, Z., Li, W., Nie, | Lu, & Zhao, K. (2024). An easy-to-follow handbook for electroencephalogram data analysis with Python. *Brain-X*, *2*(2), e64.

Lu, Z., Shafer-Skelton, A., & Golomb, J. (2022). Gaze-centered spatial representations in human hippocampus. *2022 Conference on Cognitive Computational Neuroscience*, 614–616.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *Vol. 10*, 363–401.

Neri, P., Bridge, H., & Heeger, D. J. (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. *Journal of Neurophysiology*, *92*(3), 1880–1891.

Parslow, D. M., Rose, D., Brooks, B., Fleminger, S., Gray, J. A., Giampietro, V., Brammer, M. J., Williams, S., Gasston, D., Andrew, C., Vythelingum, G. N., Ioannou, G., Simmons, A., & Morris, R. G. (2004). Allocentric spatial memory activation of the hippocampal

formation measured with fMRI. *Neuropsychology*, *18*(3), 450–461.

Rolls, E. T. (2020). Spatial coordinate transforms linking the allocentric hippocampal and egocentric parietal primate brain systems for memory, action in space, and navigation. *Hippocampus*, *30*(4), 332–353.

Roth, Z. N. (2016). Functional MRI Representational Similarity Analysis Reveals a Dissociation between Discriminative and Relative Location Information in the Human Visual System. *Frontiers in Integrative Neuroscience*, *10*, 181016.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4447–4452.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic Resonance Imaging. *Science*, *268*(5212), 889–893.

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*, *13*(11), 488–495.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging. In *Thieme*. New York: Thieme Medical Publishers. http://scholar.google.pl/scholar?hl=pl&as_sdt=0,5&q=Talairach+and+Tournoux+1988+stuttgart#4

Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences*, *2*(5), 174–183.

Tootell, R. B. H., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 811–817.

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I., & Dale, A. M. (1997). Functional Analysis of V3A and Related Areas in Human Visual Cortex. *Journal of Neuroscience*, *17*(18), 7060–7078.

Tu, Y., Li, X., Lu, Z. L., & Wang, Y. (2022). Protocol for topology-preserving smoothing of BOLD fMRI retinotopic maps of the human visual cortex. *STAR Protocols*, *3*(3), 101614.

Uchimura, M., Nakano, T., Morito, Y., Ando, H., & Kitazawa, S. (2015). Automatic representation of a visual stimulus relative to a background in the right precuneus. *European Journal of Neuroscience*, *42*(1), 1651–1659.

Uka, T., & DeAngelis, G. C. (2006). Linking neural representation to function in stereoscopic depth perception: Roles of the middle temporal area in coarse versus fine disparity discrimination. *Journal of Neuroscience*, *26*(25), 6791–6802.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383.

Wang, L., Zhu, S. J., Wang, H. Y., Qu, S. N., Zhang, Y. L., Zhang, J. H., Chen, Q. D., Xu, H. L., Han, W., Yang, B., & Sun, H. B. (2014). Common origin of green luminescence in carbon nanodots and graphene quantum dots. *ACS Nano*, *8*(3), 2541–2547.

Warnking, J., Dojat, M., Guérin-Dugué, A., Delon-Martin, C., Olympieff, S., Richard, N., Chéhikian, A., & Segebarth, C. (2002). fMRI Retinotopic Mapping—Step by Step. *NeuroImage*, *17*(4), 1665–1683.

Welchman, A. E. (2016). The Human Brain in Depth: How We See in 3D. *Annual Review of Vision Science*, *2*, 345–376.

Wilkinson, F., James, T. W., Wilson, H. R., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology*, *10*(22), 1455–1458.

Zaehle, T., Jordan, K., Wüstenberg, T., Baudewig, J., Dechent, P., & Mast, F. W. (2007). The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research*, *1137*(1), 92–103.

Supplementary

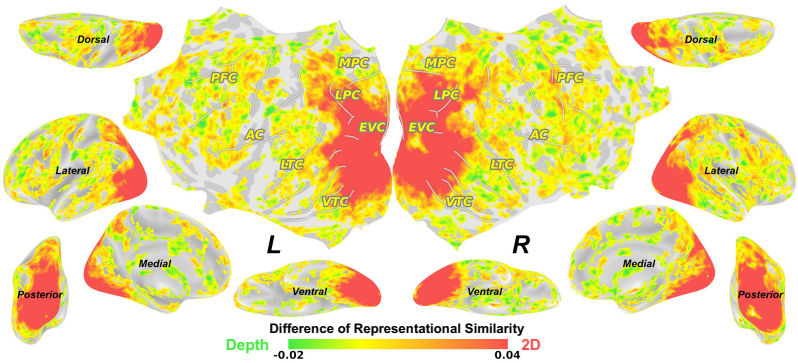

Figure S1. 2D minus depth representations from correlation-based RSA results.
Additional contrast analysis between 2D (averaging *x* and *y*) and depth (*z*) based on correlation-based RSA results replicated the finding of the spatial transition from 2D to depth along the visual hierarchy in (Finlayson et al., 2017).

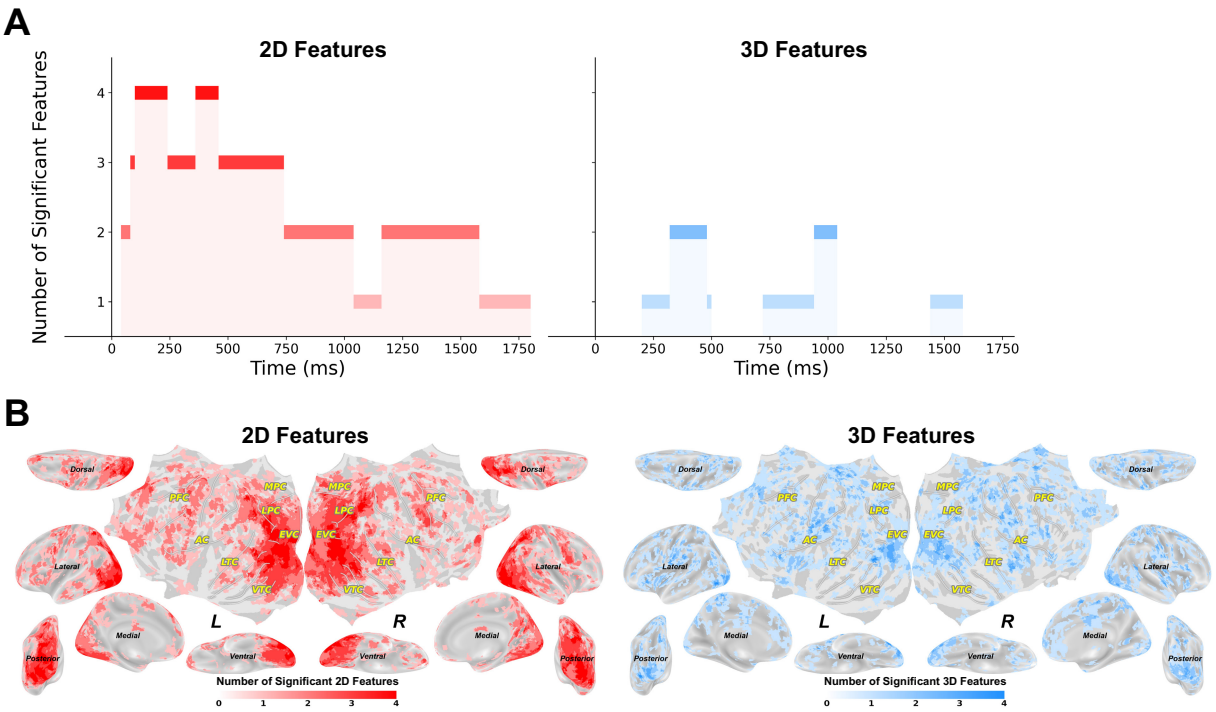

Figure S2. Number of significant 2D or 3D spatial features.
Number of significant 2D or 3D spatial features (A) over time from EEG and (B) across brain space.

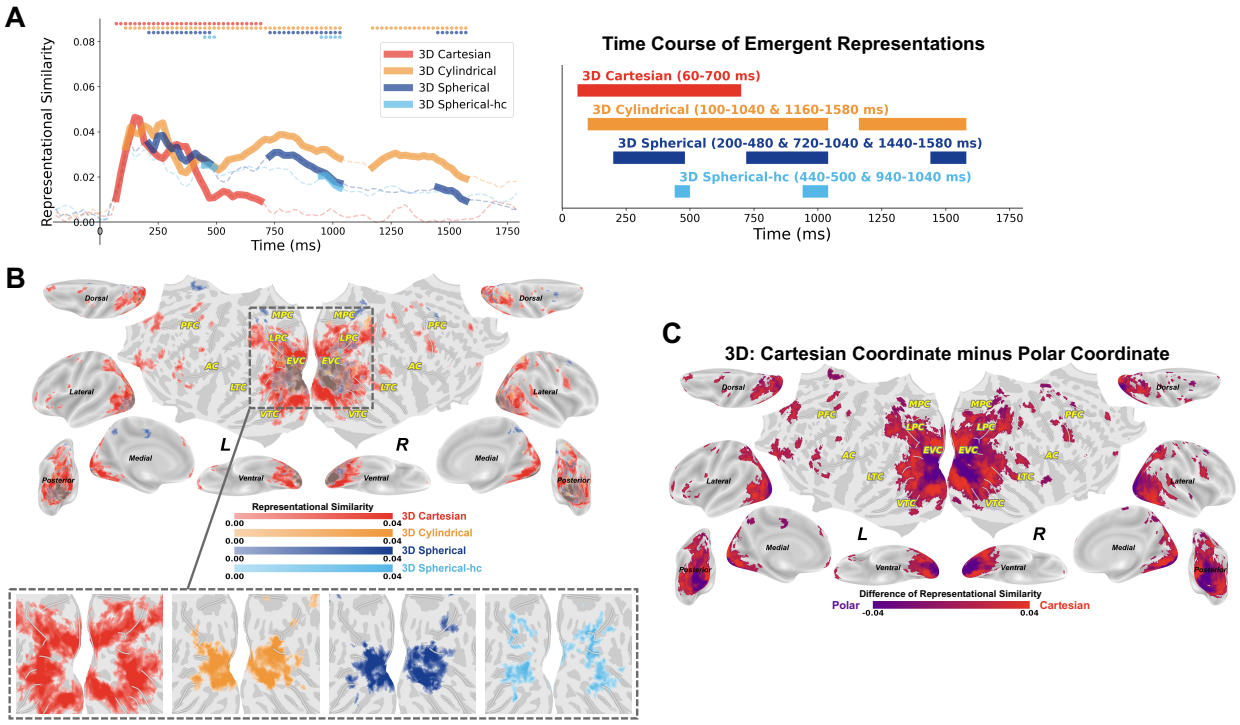

Figure S3. Emergent representations of 3D coordinate systems.

As a less stringent version of the "complete" representations of 3D coordinate systems (Figure 5), the representation of a coordinate system was considered "emergent" if (1) the average representation similarity across its internal features significantly exceeded zero, and (2) at lease one coordinate-specific feature – i.e., a feature uniquely associated with that 3D coordinate system – was also significantly represented. This emergent representation indicates that the brain shows a trendency to represent a certain coordinate system and processes at least part of that coordinate system. For instance, *r* is specific to the 3D Cylindrical coordinate system, while *x* and *y* are unique to the 3D Cartesian system.

(A) Left: Temporally significant emergent representation of 3D coordinates; Right: Time course of the significantly emergent representation for each 3D coordinate system. Thinkened lines and colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$). (B) Searchlight maps of emergent coordinate representations (permutation test, cluster-based corrected, $p < .01$). (C) Differential maps showing Cartesian minus Polar representations for 3D coordinates (permutation test, cluster-based corrected, $p < .01$). Here, we computed voxelwise differences between 3D Cartesian and 3D Polar representations. (We created a unified 3D Polar map by taking the voxelwise maximum of the emergent representations across 3D Cylindrial, 3D Spherical, and 3D Spherical (head-centered) systems, and subtracted this from 3D Cartesian map.) The resulting contrast maps revealed dissociated neural organizations of Cartesian and Polar representations.

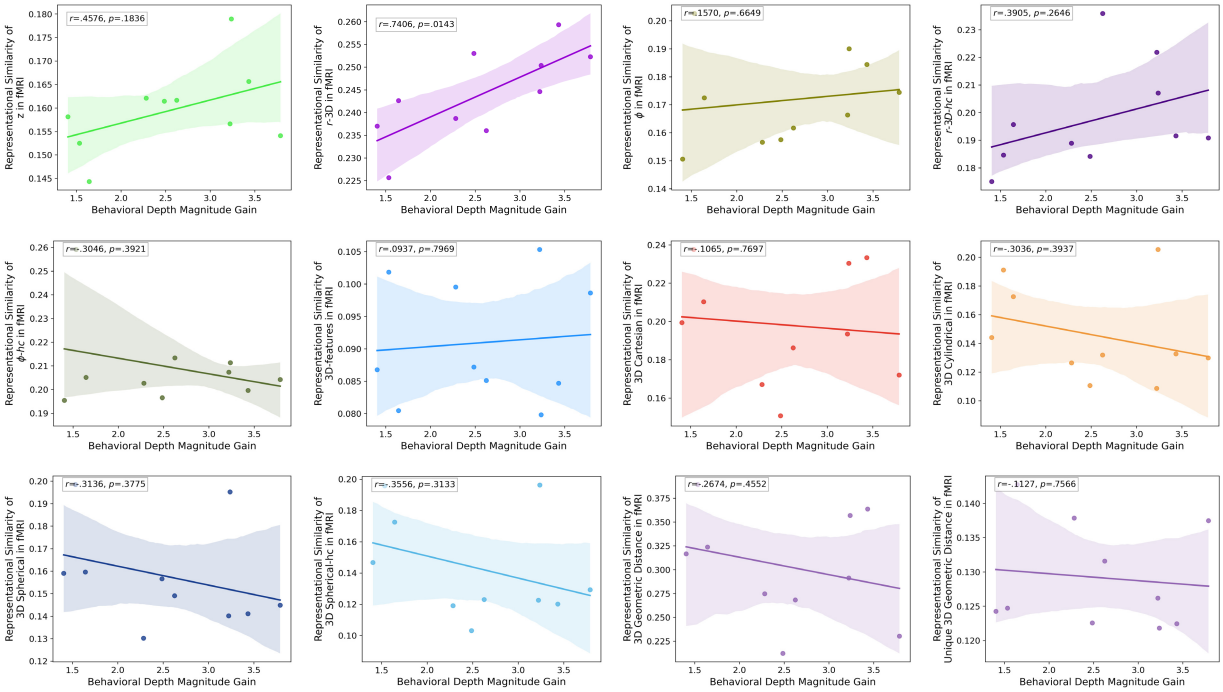

Figure S4 Correlation between behavioral depth magnitude gain and neural encoding in fMRI. Each dot represents one participant. Shaded area indicates 95% confidence interval for the fitted regression line.

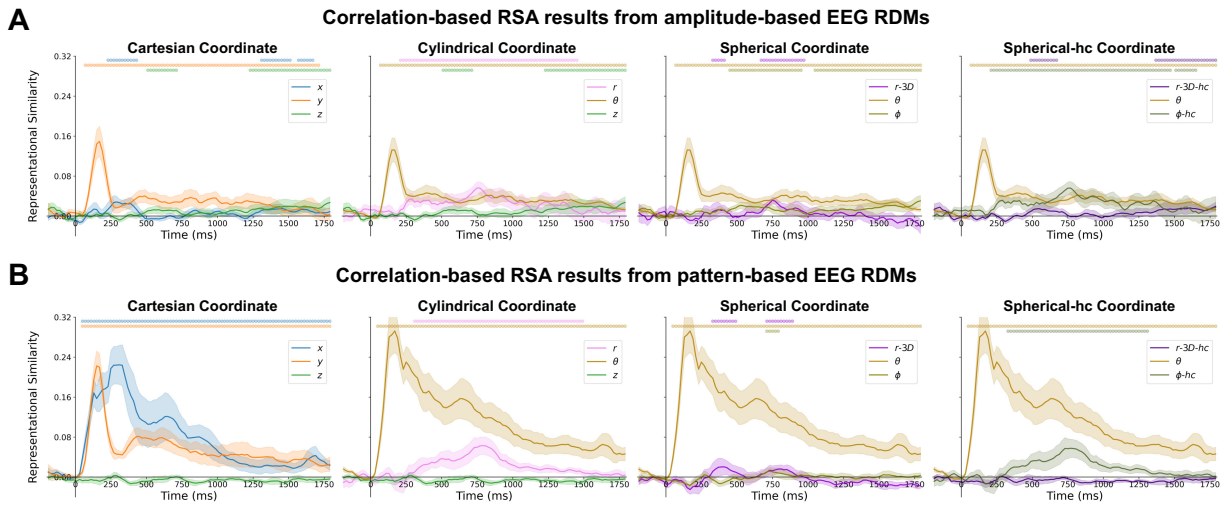

Figure S5. Correlation-based EEG RSA results of spatial features without controlling for shared variances from amplitude- and pattern-based EEG RDMs.

Time-resolved representational similarity (Spearman correlation) results between EEG temporal (A) amplitude- or (B) pattern-based RDMs and nine hypothesis-based feature RDMs, grouped by spatial coordinate systems: Cartesian (*x*, *y*, *z*), Cylindrical (*r*, *θ*, *z*), Spherical (*r-3D*, *θ*, *Φ*), and Spherical-hc (*r-3D-hc*, *θ*, *Φ-hc*). Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$).

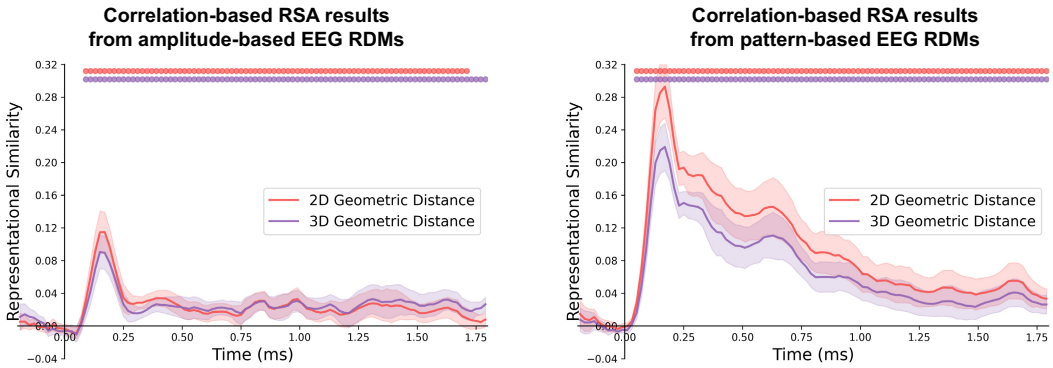

Figure S6. Correlation-based EEG RSA results of geometric distance representation without controlling for shared variances from amplitude- and pattern-based EEG RDMs.
Time-resolved representational similarity (Spearman correlation) results between EEG temporal amplitude- or pattern-based RDMs and two hypothesis-based geometric distance RDMs. Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$).

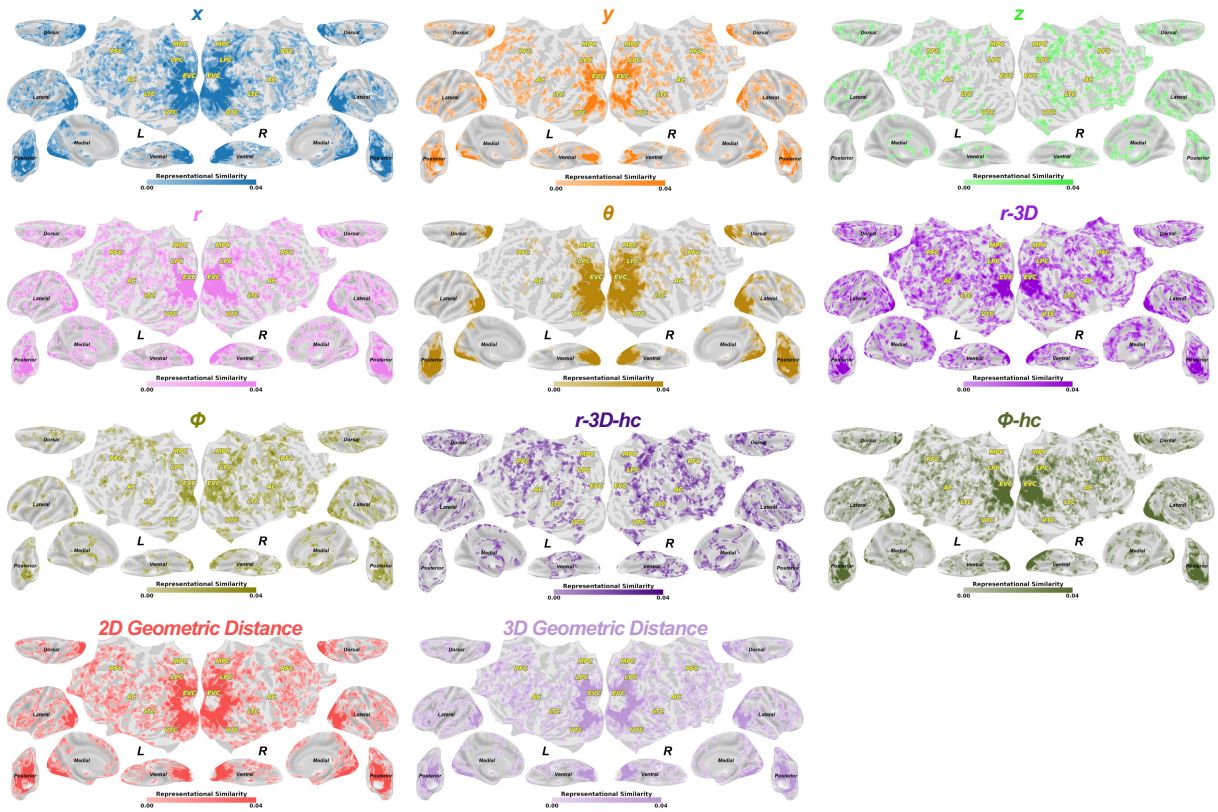

Figure S7. Correlation-based fMRI RSA results without controlling for shared variances from amplitude-based fMRI RDMs.
Searchlight representational similarity (Spearman correlation) results between fMRI searchlight amplitude-based RDMs and eleven hypothesis-based RDMs (permutation test, cluster-based corrected, $p < .01$).

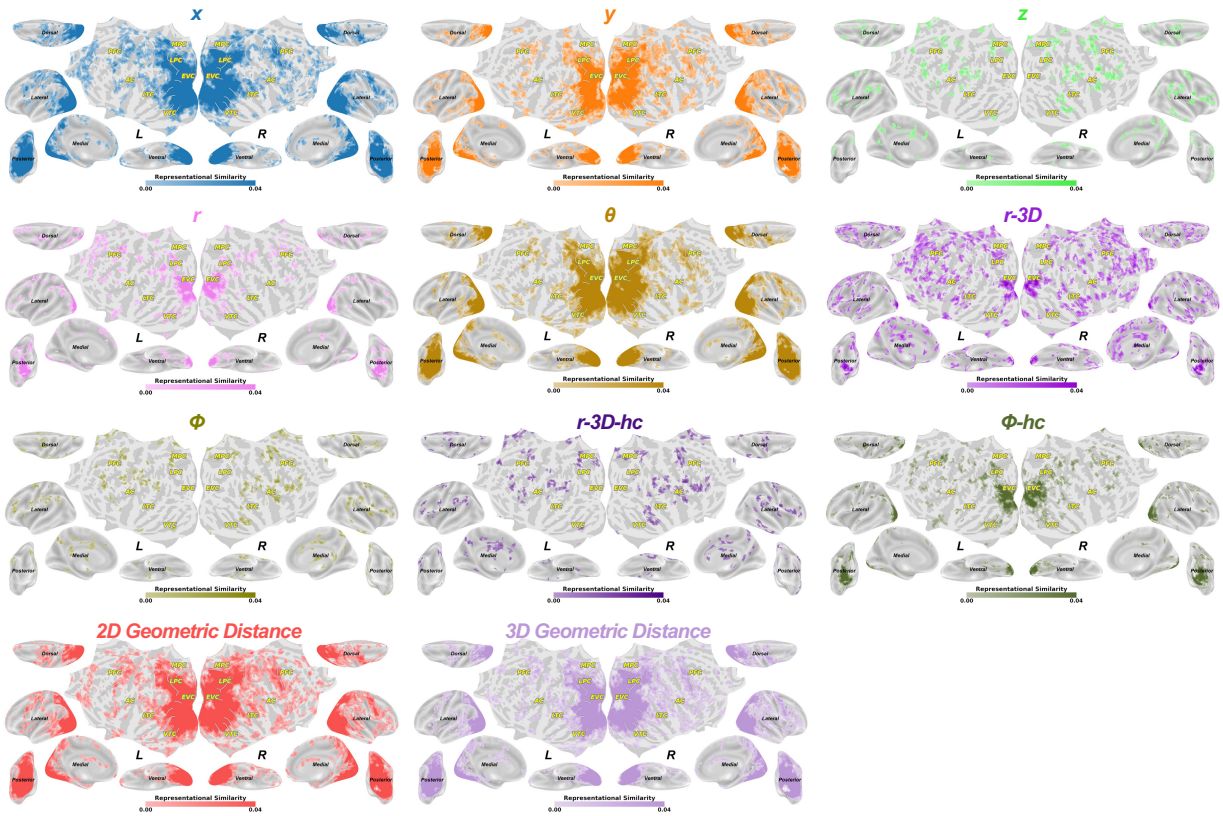

Figure S8. Correlation-based fMRI RSA results without controlling for shared variances from pattern-based fMRI RDMs.
Searchlight representational similarity (Spearman correlation) results between fMRI searchlight pattern-based RDMs and eleven hypothesis-based RDMs (permutation test, cluster-based corrected, $p < .01$).

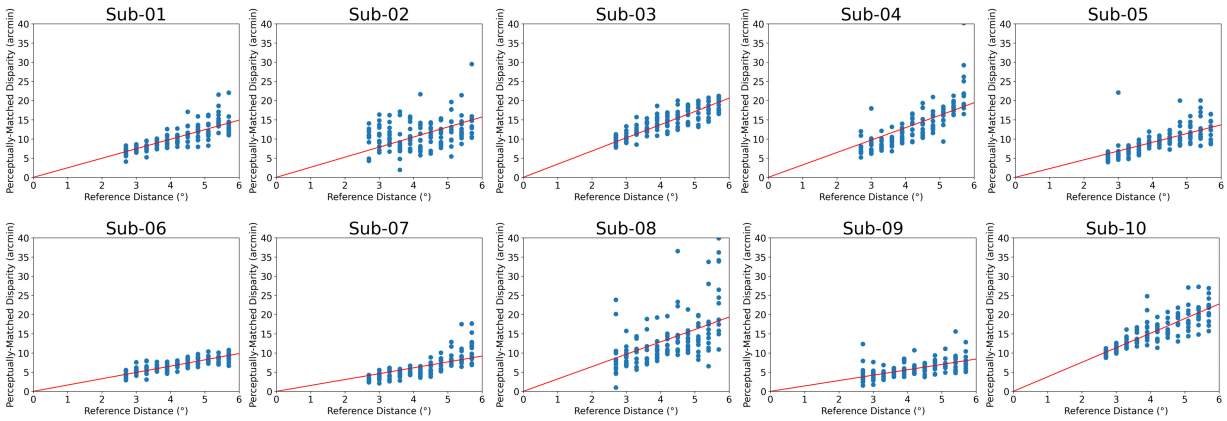

Figure S9. Individual disparity-depth functions derived from the Cube Adjustment Task.
Each panel shows data from one participant (Sub-01 to Sub-10), plotting the relationship between reference distance (in degrees of visual angle) and the perceptual matched binocular disparity (in arcmin) required to perceive a cube with equal 2D and depth dimensions. Each blue dot represents a single trial in which participants adjusted the relative front-back disparity of four RDS patches to form a perceptually cube shape. Red lines indicate the best-fitting linear function (intercept fixed at zero) obtained via least squares regression. These individualized functions were used to convert 3D spatial positions into participant-specific disparity values in the EEG and fMRI main task.

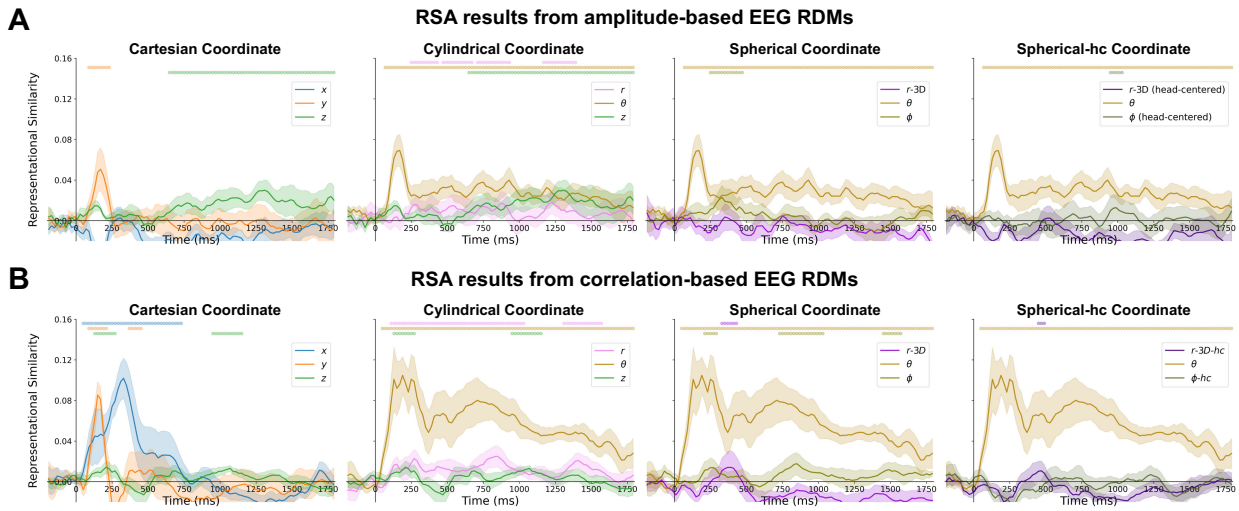

Figure S10. Partial correlation-based EEG RSA results of spatial features from amplitude- and pattern-based EEG RDMs.

Time-resolved representational similarity (partial Spearman correlation) results between EEG temporal (A) amplitude- or (B) correlation-based RDMs and nine hypothesis-based feature RDMs, grouped by spatial coordinate systems: Cartesian (x, y, z), Cylindrical (r, θ, z), Spherical (r-$3D$, θ, Φ), and Spherical-hc (r-$3D$-hc, θ, Φ-hc). Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$).

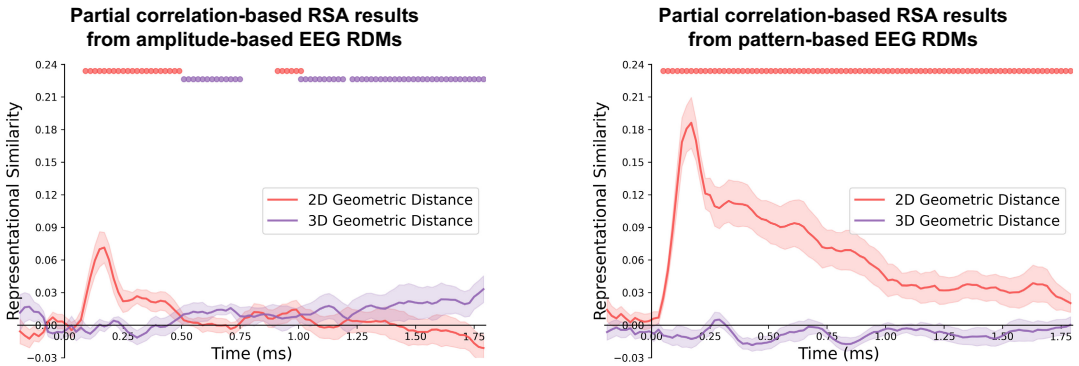

Figure S11. Partial correlation-based EEG RSA results of geometric representation from amplitude- and pattern-based EEG RDMs.

Time-resolved representational similarity (partial Spearman correlation) results between EEG temporal amplitude- or correlation-based RDMs and two geometric distance RDMs. Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$).

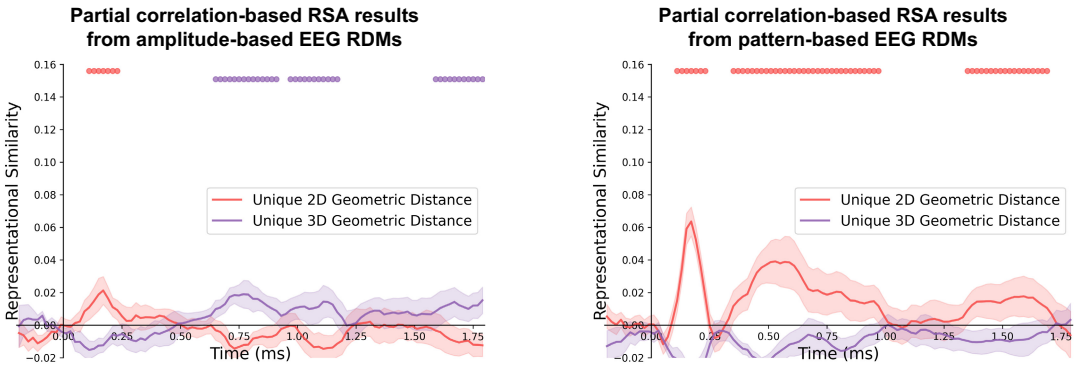

Figure S12. Partial correlation-based EEG RSA results of unique geometric representation from amplitude- and pattern-based EEG RDMs.
Time-resolved representational similarity (partial Spearman correlation, also controlling nine feature RDMs) results between EEG temporal amplitude- or correlation-based RDMs and two geometric distance RDMs. Shaded areas indicate ±1 SEM across participants. Colored dots indicate significant timepoints (permutation test, cluster-corrected, $p < .01$).

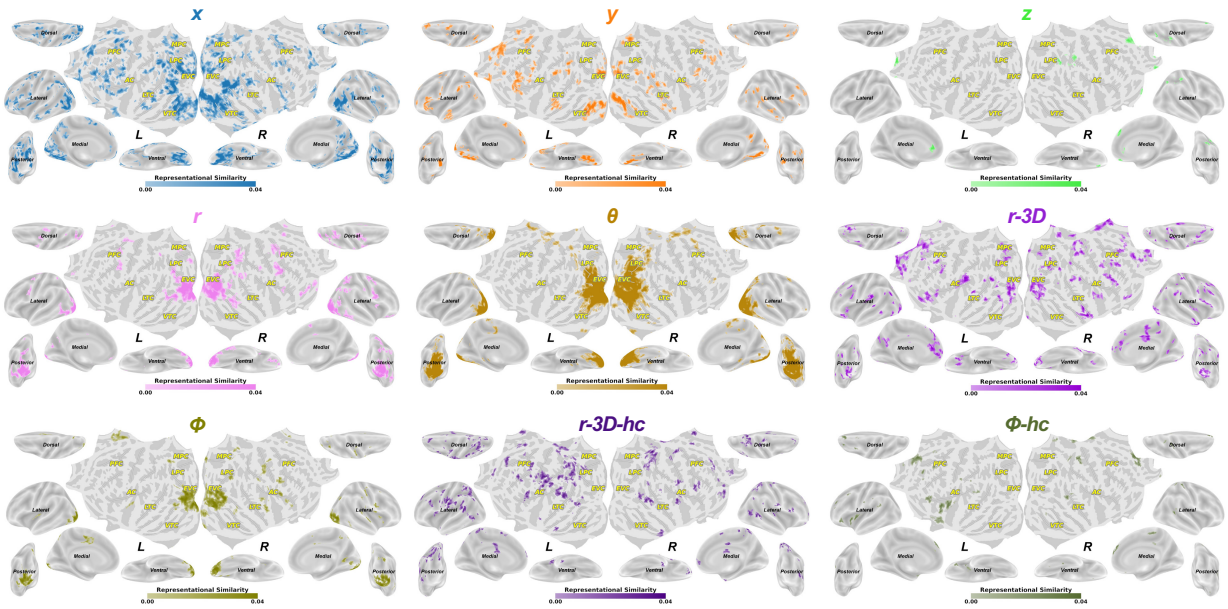

Figure S13. Partial correlation-based fMRI RSA results of spatial features from amplitude-based fMRI RDMs.
Searchlight representational similarity (partial Spearman correlation) results between fMRI searchlight amplitude-based RDMs and nine hypothesis-based feature RDMs (permutation test, cluster-based corrected, $p < .01$).

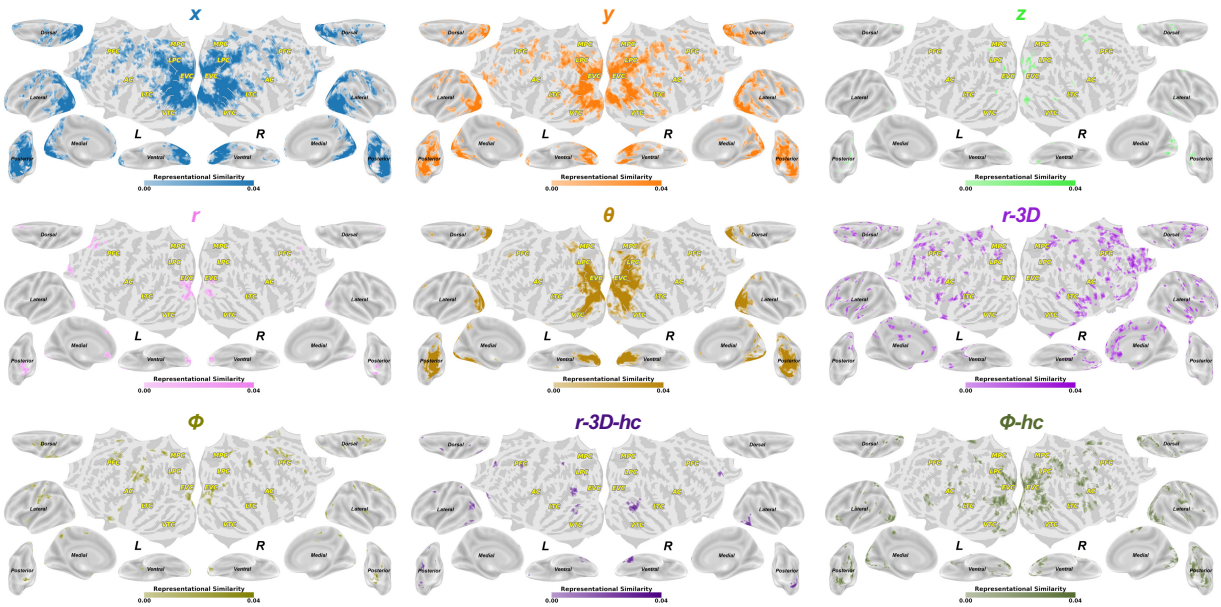

Figure S14. Partial correlation-based fMRI RSA results of spatial features from pattern-based fMRI RDMs.
Searchlight representational similarity (partial Spearman correlation) results between fMRI searchlight pattern-based RDMs and nine hypothesis-based feature RDMs (permutation test, cluster-based corrected, $p < .01$).

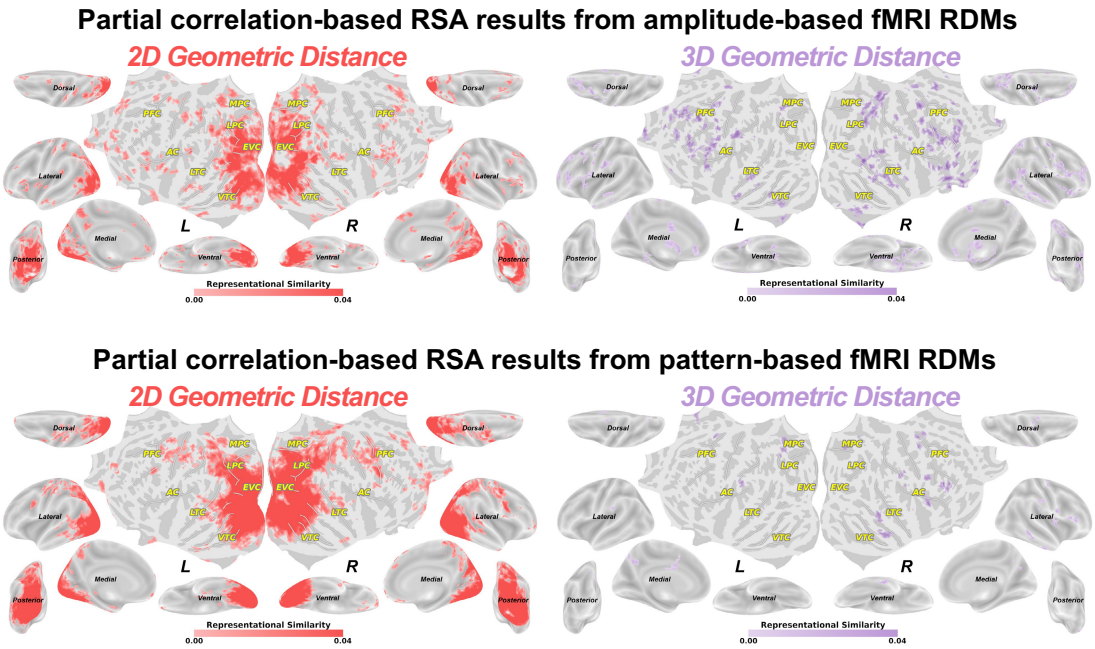

Figure S15. Partial correlation-based fMRI RSA results of geometric distance representation from amplitude- and pattern-based fMRI RDMs.
Searchlight representational similarity (partial Spearman correlation) results between fMRI searchlight pattern-based RDMs and two hypothesis-based geometric distance RDMs (permutation test, cluster-based corrected, $p < .01$).

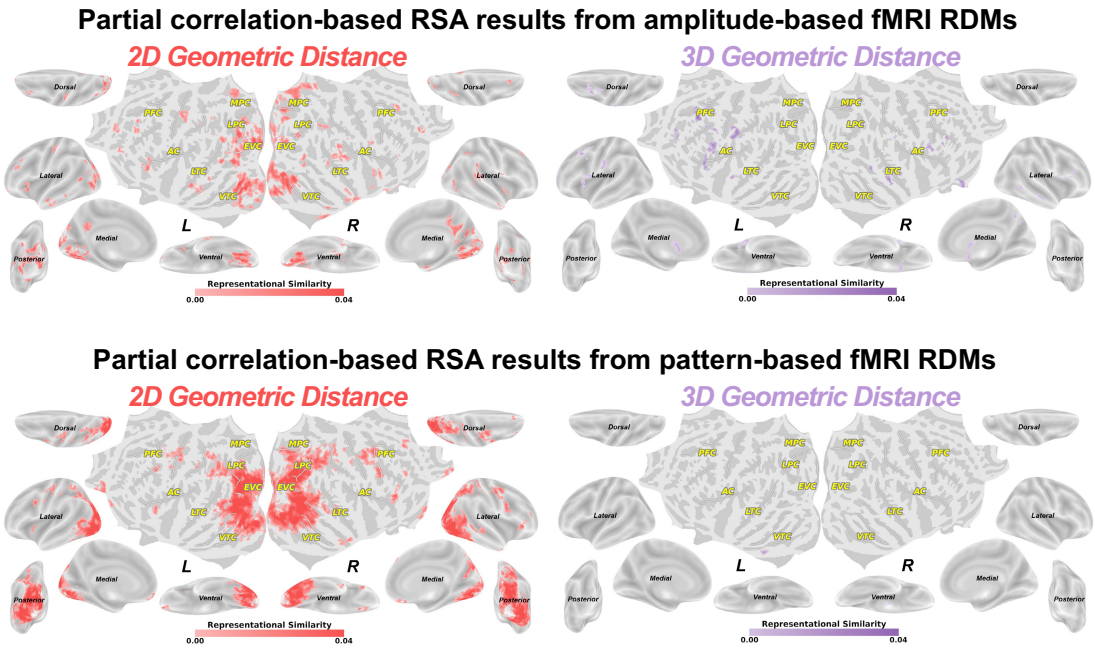

Figure S16. Partial correlation-based fMRI RSA results of unique geometric distance representation from amplitude- and pattern-based fMRI RDMs.
Searchlight representational similarity (partial Spearman correlation, also controlling nine feature RDMs) results between fMRI searchlight pattern-based RDMs and two hypothesis-based geometric distance RDMs (permutation test, cluster-based corrected, $p < .01$).

(See the video via the link:
https://github.com/ZitongLu1996/3D_Visual_Perception/blob/master/video.mp4)

Video S1. Visualization of combined EEG and fMRI RSA results of spatial feature representations. In this video, if the EEG RSA result of a certain spatial feature is significant, we highlighted the corresponding significant fMRI RSA result of the feature on inflated cortical surfaces (based on Figure 3A and 4A).